Efficient One-pass Multi-view Subspace Clustering with Consensus Anchors

Suyuan Liu[1,†], Siwei Wang[1,†], Pei Zhang[1], Xinwang Liu[1,*], Kai Xu[1], Changwang Zhang[2] and Feng Gao[3]

[1] School of Computer, National University of Defense Technology, Changsha, China, 410073
[2] Technology and Engineering Group, Tencent Technology (Shenzhen) Co., Ltd., Shenzhen, China, 518064
[3] School of Arts, Peking University, Beijing, China, 100871

Abstract

Multi-view subspace clustering (MVSC) optimally integrates multiple graph structure information to improve clustering performance. Recently, many anchor-based variants are proposed to reduce the computational complexity of MVSC. Though achieving considerable acceleration, we observe that most of them adopt fixed anchor points separating from the subsequential anchor graph construction, which may adversely affect the clustering performance. In addition, post-processing is required to generate discrete clustering labels with additional time consumption. To address these issues, we propose a scalable and parameter-free multi-view subspace clustering method to directly output the clustering labels with optimal anchor graph, termed as Efficient One-pass Multi-view Subspace Clustering with Consensus Anchors (EOMSC-CA). Specially, we integrate the anchor points selection and graph construction into a unified optimization framework to boost clustering performance. Meanwhile, by imposing a graph connectivity constraint, our algorithm directly outputs the clustering labels without any post-processing procedures as previous methods do. The proposed approach is proven to has linear time complexity respecting to the data size. Extensive experiments on various benchmark datasets verify the superiority of the proposed method compared to the existing state-of-the-art multi-view subspace clustering competitors over the effectiveness and efficiency. Our code is publicly available at https://github.com/Tracesource/EOMSC-CA.

Introduction

In the era of big data, datasets are high dimensional, but their inherent dimension is often much smaller than the dimension of the ambient space. In fact, many high-dimensional data can be represented by samples drawn from the union of low-dimensional subspaces. Based on the above assumptions, the goal of subspace clustering is to cluster the data lied in the linear subspaces and determine the low-dimensional subspace corresponding to each cluster (Vidal 2011; Elhamifar and Vidal 2013). However, data usually are collected from different sources or diverse domains in many real-world problems (Wang et al. 2019; Zhang et al. 2020; Liu et al. 2021b; Wang et al. 2021). For instance, the same news can be reported in the form of text, images and videos in different languages (Kang et al. 2019). Therefore, in order to make full use of multi-view information, many multi-view subspace clustering methods have been proposed in recent years to explore consistency and complementarity across views (Tang et al. 2018; Wang et al. 2020; Liu et al. 2021a; Zhang et al. 2021; Zhou et al. 2020a). Most of the existing multi-view clustering methods use graph-based models, which first learn a common similarity matrix from the features of different views and then execute spectral clustering algorithm (Ng, Jordan, and Weiss 2002) to obtain the final clustering results (Liu et al. 2012; Ding and Fu 2014; Zhan et al. 2017; Xu et al. 2018; Zhou et al. 2021). Existing multi-view subspace clustering algorithms suffer from high time consumption, making it difficult to process large-scale data in reality. The high time complexity mainly comes from the construction of the similarity matrix, the calculation of the spectral embedding in spectral clustering, and the discretization of the spectral embedding to obtain the clustering results.

In order to improve efficiency, it has recently been proposed to apply bipartite graph learning, using the relationship between anchor points and data points to represent the relationship between all data points (Chen and Cai 2011; Adler, Elad, and Hel-Or 2015; Li et al. 2020; Liu et al. 2021c). Then the graph size of each view is reduced from $n \times n$ to $n \times m$ where m denotes the number of anchor points. However, existing anchor point selection applies heuristic sampling strategies, such as k-means or random sampling, which prevents the mutual negotiation between anchor point selection and graph construction to achieve optimal clustering. Besides, existing graph-based methods usually consist of two stages. First, a graph is constructed from raw data, and then perform post-processing on the fixed graph to obtain the clustering result(Zhou et al. 2020b; Gao et al. 2020; Liu et al. 2019). The final cluster structure is not clearly shown in the graph constructed in the previous stage, and the clustering performance heavily dependents on the quality of graph (Nie et al. 2016b, 2017). Moreover, multi-view clustering algorithms usually encounter a large number of hyper-parameters. How to determine these hyper-parameters which severely reduce the effectiveness of algorithms remains to be a challenge.

In this paper, we propose a novel scalable multi-view

*Corresponding Author

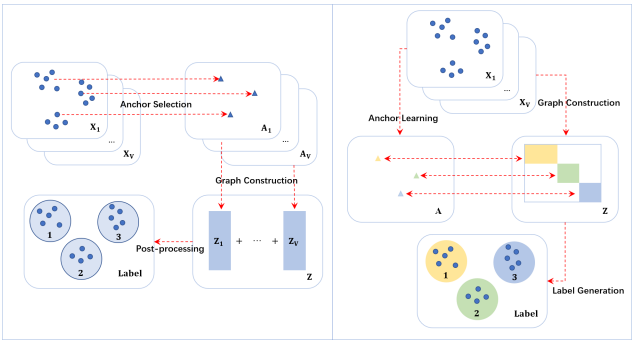


Figure 1: The framework of existing anchor-based methods(left) and our method(right). The existing methods first select anchor points from the data point of each view according to the sampling strategy, then construct the graph representation separately, and finally perform post-processing to obtain the final cluster label. In contrast, our method directly learns consistent anchor points and common graph representations from multi-view data. The selection of anchor points and the construction of the graph influence each other, and finally clusters label are directly generated through graphs with exactly k connected components.

subspace clustering method termed as Efficient One-pass Multi-view Subspace Clustering with Consensus Anchors (EOMSC-CA) to directly output the clustering labels without post-processing procedure. Firstly, we integrate anchor learning and fused graph construction into a unified optimization framework, and the two processes can be negotiated with each other to improve clustering performance. At the same time, we adaptively learn view weights in the framework without additional hyper-parameters to balance the effects of the view. In addition, we impose rank constraints on the Laplacian graph of the similarity matrix so that the final joint graph has exactly several connected components of clusters to indicate each cluster. Extensive experiments on various large-scale benchmark datasets verify the effectiveness and efficiency of our method.

The contributions of this paper can be summarized as follows:

- Instead of existing fixed sampling anchors, we unify anchor learning and fused graph construction into a unified and flexible framework so that the two processes seamlessly contribute to each other and boost clustering performance.
- We propose a novel scalable multi-view subspace clustering method to output the optimal anchor graph with exactly k-connected components of the clusters, which can be directly used for clustering without performing discretization procedures necessary for other graph-based clustering methods.
- Experimental results on multiple benchmark datasets verify the effectiveness and efficiency of the proposed algorithm, especially on large-scale multi-view datasets by large margins. Our method is proven to be linearly time complexity respecting to sample numbers and does not involve any hyper-parameters. These clearly make the proposed EOMSC-CA more suitable for large-scale multi-view data clustering tasks.

Table 1: Main notations used throughout the paper.

Notion	Definition
n	Number of data points
k	Number of clusters
v	Number of views
m	Number of anchor points
l	Dimension of anchor matrix
d_p	Dimension of the p-th view
d	$\sum_{p=1}^{v} d_i$
$\boldsymbol{\beta}$	View coefficient
$\mathbf{X}_p \in \mathbb{R}^{d_p \times n}$	Data matrix of the p-th view
$\mathbf{W}_p \in \mathbb{R}^{d_p \times l}$	Projection matrix of the p-th view
$\mathbf{A} \in \mathbb{R}^{l \times m}$	Consensus anchor matrix
$\mathbf{Z} \in \mathbb{R}^{m \times n}$	Fused anchor graph

Background

In this section, we provide an overview to the subspace clustering and multi-view subspace clustering, and then introduce the previous works to our research. Table 1 summarizes some notations used throughout the paper.

Subspace Clustering

Given data $\mathbf{X} \in \mathbb{R}^{d \times n}$ with d dimension and n samples, subspace clustering assumes that each data point can be expressed by a linear combination of other data points in the same subspace. The first step of subspace clustering is to construct a graph, which solves the following problem:

$$\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{X}\mathbf{S}\|_F^2 + \lambda f(\mathbf{S}), s.t. \mathbf{S} \geq 0, \mathbf{S}^\top \mathbf{1} = \mathbf{1}, \quad (1)$$

where $\mathbf{S} \in \mathbb{R}^{n \times n}$ denotes the non-negative self-representation matrix, $f(\cdot)$ represents the regularization functions, and λ is a balance parameter. The $\mathbf{S}^\top \mathbf{1} = \mathbf{1}$ ensures that each column of $\mathbf{S}$ adds up to 1. With the size of $\mathbf{S}$, the graph construction in the first stage often takes at least $\mathcal{O}\left(n^3\right)$.

After obtaining the coefficient matrix $\mathbf{S}$, the second step is to get the spectral embedding $\mathbf{F} \in \mathbb{R}^{n \times k}$ with the symmetric similarity matrix $\mathbf{W}$ constructed as $\mathbf{W} = \frac{\mathbf{S}+\mathbf{S}^\top}{2}$,

$$\min_{\mathbf{F}} \operatorname{Tr}\left(\mathbf{F}^\top \mathbf{L}\mathbf{F}\right), s.t. \mathbf{F}^\top \mathbf{F} = \mathbf{I}_k, \quad (2)$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian with diagonal matrix $\mathbf{D}$ defined as $d_{ii} = \sum_{j=1}^{n} s_{ij}$, and F is the spectral embedding which can get the clustering result after k-means algorithm. The spectral clustering stage takes $\mathcal{O}\left(n^3\right)$ complexity.

Multi-view Subspace Clustering

Multi-view subspace clustering is developed on the basis of subspace clustering. For multi-view data $\{\mathbf{X}_p\}_{p=1}^{v}$, where

$\mathbf{X}_p \in \mathbb{R}^{d_p \times n}$ represents the p-th view data with d_p dimensions, the following model is solved:

$$\min_{\mathbf{S}_p, \mathbf{S}} \|\mathbf{X}_p - \mathbf{X}_p \mathbf{S}_p\|_F^2 + \lambda f(\mathbf{S}, \mathbf{S}_p), s.t. \mathbf{S}_p \geq 0, \mathbf{S}_p^\top \mathbf{1} = \mathbf{1}, \tag{3}$$

where $f(\cdot)$ represents the regularization term which aims to get a consensus graph among various views. Performing spectral clustering on the fused global graph $\mathbf{Z}$ and then the clustering results can be reached.

Based on the above framework, many multi-view subspace clustering methods have recently been proposed. Gao et al. (2015) propose to learn an independent subspace representation in each view, and then use a unified index matrix to get a common clustering result. In order to explore the complementarity of multi-view representations, Cao et al. (2015) propose to introduce Hilbert-Schmidt independence criterion as a regularization term. Considering that the subspace representation learned from multiple views may be redundant, Zhang et al. (2015) regard the subspace representation matrix of different views as a tensor, and explore the intersection information among views with low rank constraints. Another important point is to impose appropriate constraints on the subspace representation, such as low-rank constraints (Wang et al. 2018) or sparse constraints (Lu, Yan, and Lin 2016). Brbić and Kopriva (2018) suggest to build a consensus affinity matrix that satisfies both low-rank and sparsity.

However, the above methods poorly handle with scalability problem. Specifically, in addition to fusing multi-view information, multi-view subspace clustering requires graph construction and spectral embedding as the same as single-view.

Anchor-based Multi-view Clustering

Anchor graph is recently applied to solve large-scale data clustering problems (Kang et al. 2020; Li et al. 2020; Ou et al. 2020; Kang et al. 2021; Sun et al. 2021). The principle of anchor graph is to select a part of the representative points from the original data called anchor points, and then use the relationship between anchor points and the entire data points to cover the complete affinity. Specifically, a small graph $\mathbf{Z} \in \mathbb{R}^{m \times n}$ is constructed to replace the original graph $\mathbf{S} \in \mathbb{R}^{n \times n}$. With the application of anchor points, the complexity of subspace clustering can be greatly reduced while maintaining considerable efficiency.

Some recent works have applied anchor graphs into multi-view clustering. Kang et al. (2020) propose to get anchors of each view with k-means at first, and then construct anchor graphs separately. Li et al. (2020) provide an anchor selection strategy based to importance build anchor graphs independently in each view and then fuse them into a consensus one.

Although anchor-based multi-view clustering methods achieve the state-of-the-art performance, there still exists some limitations. For one thing, the selection of anchor points is separated in each view without mutual negotiation with the clustering process. For another, inevitable hyper-parameters affect their efficiency on large-scale data. Moreover, existing works need to preform a post-processing process to get the final result, and the cluster structure is not obvious in the constructed graph. To solve these problems, we propose a novel method named EOMSC-CA in the next section.

The Proposed Methodology

In this section, we will describe in detail the novel subspace clustering method termed as EOMSC-CA, including the motivation, formulation, optimization, and complexity analysis.

Motivation

Reducing the redundancy of data in large-scale data clustering is the key to improve efficiency. Actually, a small number of instances are enough to reconstruct the underlying subspaces. Therefore, the existing works propose to select some points from the original data as anchor points to reconstruct the relationship structure.

However, the existing anchor-based multi-view subspace methods all perform heuristic sampling strategies, which means that the selection of anchor points and the construction of the graph are separated. After selecting the anchor points separately in each view, graphs are then constructed as follow,

$$\min_{\mathbf{Z}_p} \sum_{p=1}^{v} \left\| \mathbf{X}_p - \mathbf{A}_p \left(\mathbf{Z}_p\right)^\top \right\|_F^2, s.t. \mathbf{Z}_p \geq 0, \mathbf{Z}_p^\top \mathbf{1} = \mathbf{1}. \tag{4}$$

where $\mathbf{A}_p$ and $\mathbf{Z}_p$ are anchor matrix and anchor graph of the p-th view, respectively.

The graph constructed in each view cannot integrate the complementary information of multiple views well, which need to perform fusion algorithm to get a consensus graph. Besides, after constructing the graph with subspace clustering, it is usually necessary to perform spectral clustering to obtain the spectral embedding, and then adopts k-means to obtain the clustering results. This two-stage process causes the final cluster structure unclear in the constructed graph, and the clustering result is highly dependent on the quality of the graph (Nie et al. 2016b).

Formulation of Problem

To address the above challenges, we aim to **learn** anchors through optimization instead of sampling. Based on the assumption that high-dimensional data in different views share a consensus low-dimensional subspace, the learned anchors should be consistent in the consensus subspace. Define the projection matrix $\{\mathbf{W}_p\}_{p=1}^v$, we can align the consensus anchors $\mathbf{A}$ with the original data of the p-th view.

In order to make the obtained graph a clear clustering structure and directly obtain the cluster labels without post-processing, we hope that the graph has exactly k connected components. For the graph $\mathbf{Z} \in \mathbb{R}^{m \times n}$ constructed based on the anchor method, we define the augmented graph $\mathbf{S}$ as

$$\mathbf{S} = \begin{bmatrix} & \mathbf{Z}^\top \\ \mathbf{Z} & \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}. \tag{5}$$

Then we get the normalized Laplacian matrix $\widetilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{S} \mathbf{D}^{-\frac{1}{2}}$, where $\mathbf{D}$ is a diagonal matrix with i-th element

computed by $d_i = \sum_{j=1}^{n+m} s_{ij}$. According to the following Theorem, the rank constraint of Laplacian matrix ensures $\mathbf{S}$ to be k-connected.

Theorem 1 : The number of connected components in $\mathbf{S}$ equals to the cardinality of the eigenvalue zero of the normalized Laplacian matrix $\widetilde{\mathbf{L}}$.

Therefore, the graph $\mathbf{S}$ has exactly k connected components if $rank(\widetilde{\mathbf{L}}) = n + m - k$. Since $\mathbf{S}$ is composed of $\mathbf{Z}$, the number of connected components in $\mathbf{Z}$ is the same as $\mathbf{S}$.

Eventually, our proposed EOMSC-CA can be formulated as

$$\min_{\boldsymbol{\beta},\mathbf{W}_p,\mathbf{A},\mathbf{Z}} \sum_{p=1}^{v} \beta_p^2 \left\|\mathbf{X}_p - \mathbf{W}_p\mathbf{A}\mathbf{Z}\right\|_F^2, \tag{6}$$

$$s.t. \boldsymbol{\beta}^\top \mathbf{1} = 1, \mathbf{W}_p^\top \mathbf{W}_p = \mathbf{I}_l, \mathbf{A}^\top \mathbf{A} = \mathbf{I}_m,$$

$$\mathbf{Z} \geq 0, \mathbf{Z}^\top \mathbf{1} = \mathbf{1}, rank\left(\widetilde{\mathbf{L}}\right) = n + m - k,$$

where $\mathbf{X}_p \in \mathbb{R}^{d_p \times n}$ denotes the original data of the p-th view with d_p dimension, β_p is the weight coefficient that balances the influence of each view, $\mathbf{W}_p$ is the projection matrix of the p-th view and $\mathbf{A} \in \mathbb{R}^{l \times m}$ is the consensus anchor matrix with m anchors and l dimension.

Optimization

To solve the above optimization, all variables are updated alternatively, which means updating one variable with others being fixed.

Optimize Z by fixing other variables When $\mathbf{W}_p, \mathbf{A}, \boldsymbol{\beta}$ are fixed, optimizing $\mathbf{Z}$ is equal to solve the following equation:

$$\min_{\mathbf{Z}} \sum_{p=1}^{v} \beta_p^2 \operatorname{Tr}\left(\mathbf{Z}^\top\mathbf{Z} - 2\mathbf{X}_p^\top\mathbf{W}_p\mathbf{A}\mathbf{Z}\right), \tag{7}$$

$$s.t. \mathbf{Z} \geq 0, \mathbf{Z}^\top \mathbf{1} = \mathbf{1}, \operatorname{rank}\left(\widetilde{\mathbf{L}}\right) = n + m - k.$$

Define $\sigma_i\left(\widetilde{\mathbf{L}}\right)$ as the i-th smallest eigenvalue of $\widetilde{\mathbf{L}}$. Then $\sigma_i\left(\widetilde{\mathbf{L}}\right) \geq 0$ since $\widetilde{\mathbf{L}}$ is semi-definite. With λ to be large enough, Eq. (7) equals to the following problem:

$$\min_{\mathbf{Z}} \sum_{p=1}^{v} \beta_p^2 \operatorname{Tr}\left(\mathbf{Z}^\top\mathbf{Z} - 2\mathbf{X}_p^\top\mathbf{W}_p\mathbf{A}\mathbf{Z}\right) + \lambda \sum_{i=1}^{k} \sigma_i\left(\widetilde{\mathbf{L}}\right), \tag{8}$$

$$s.t. \mathbf{Z} \geq 0, \mathbf{Z}^\top \mathbf{1} = \mathbf{1}.$$

According to the Ky Fan's Theorem (Fan 1949), the Eq. (8) can be rewritten as,

$$\min_{\mathbf{Z},\mathbf{F}} \sum_{p=1}^{v} \beta_p^2 \operatorname{Tr}\left(\mathbf{Z}^\top\mathbf{Z} - 2\mathbf{X}_p^\top\mathbf{W}_p\mathbf{A}\mathbf{Z}\right) + \lambda \operatorname{Tr}\left(\mathbf{F}^\top\widetilde{\mathbf{L}}\mathbf{F}\right) \tag{9}$$

$$s.t. \mathbf{Z} \geq 0, \mathbf{Z}^\top \mathbf{1} = \mathbf{1}, \mathbf{F}^\top\mathbf{F} = \mathbf{I}_k$$

where $\mathbf{F} \in \mathbb{R}^{(n+m)\times k}$ is the indicator matrix.

When $\mathbf{Z}$ is fixed, Eq. (9) becomes:

$$\min_{\mathbf{F}^\top\mathbf{F}=\mathbf{I}_k} \operatorname{Tr}\left(\mathbf{F}^\top\widetilde{\mathbf{L}}\mathbf{F}\right). \tag{10}$$

To solve Eq. (10), we need to compute the eigenvectors of $\widetilde{\mathbf{L}}$, which takes $\mathcal{O}(k(n+m)^2)$ complexity. In order to reduce the complexity, we compute the eigenvectors of $\mathbf{Z}$ instead of $\mathbf{S}$. Specifically, $\mathbf{F}$ and $\mathbf{D}$ in Eq. (10) can be decomposed as

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{(n)} \\ \mathbf{F}_{(m)} \end{bmatrix}, \mathbf{D}_S = \begin{bmatrix} \mathbf{D}_{(n)} & \\ & \mathbf{D}_{(m)}, \end{bmatrix} \tag{11}$$

where $\mathbf{F}_{(n)}$ is the first n rows of $\mathbf{F}$ which represents the indicator of data points and $\mathbf{F}_{(m)}$ is the rest m rows of $\mathbf{F}$ which denotes the indicator of anchor points. Then we can rewritten Eq. (10) as follws:

$$\min_{\mathbf{F}_{(n)}^\top\mathbf{F}_{(n)}+\mathbf{F}_{(m)}^\top\mathbf{F}_{(m)}=\mathbf{I}_k} \operatorname{Tr}\left(\mathbf{F}_{(n)}^\top\mathbf{D}_{(n)}^{\frac{1}{2}}\mathbf{Z}^\top\mathbf{D}_{(m)}^{\frac{1}{2}}\mathbf{F}_{(m)}^\top\right). \tag{12}$$

Eq. (12) can be easily solved by the following theorem.

Theorem 2 : Suppose $\mathbf{Q} \in \mathbb{R}^{n\times m}$, $\mathbf{A} \in \mathbb{R}^{n\times k}$ and $\mathbf{B} \in \mathbb{R}^{m\times k}$, we have the follow problem

$$\max_{\mathbf{A}^T\mathbf{A}+\mathbf{B}^T\mathbf{B}=\mathbf{I}_k} \operatorname{Tr}\left(\mathbf{A}^T\mathbf{Q}\mathbf{B}\right). \tag{13}$$

The optimal solutions of the problem are $\mathbf{A} = \frac{\sqrt{2}}{2}\mathbf{U}$ and $\mathbf{B} = \frac{\sqrt{2}}{2}\mathbf{V}$, where $\mathbf{U}$ and $\mathbf{V}$ are respectively corresponding to the top k left and right singular vectors.

After optimizing $\mathbf{F}$, we optimize $\mathbf{Z}$ in Eq. (9) with the optimal $\mathbf{F}$. To solve the problem, we need to employ the following equation:

$$\operatorname{Tr}\left(\mathbf{F}^\top\widetilde{\mathbf{L}}\mathbf{F}\right) = \frac{1}{2}\sum_{i=1}^{n+m}\sum_{j=1}^{n+m}\left\|\frac{\mathbf{f}^i}{\sqrt{\mathbf{D}(i,i)}} - \frac{\mathbf{f}^j}{\sqrt{\mathbf{D}(j,j)}}\right\|_2^2 s_{ij}. \tag{14}$$

With the definition in Eq. (11), we can further transform the above formula as follows,

$$\operatorname{Tr}\left(\mathbf{F}^\top\widetilde{\mathbf{L}}\mathbf{F}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\frac{\mathbf{f}_{(n)}^i}{\sqrt{\mathbf{D}_{(n)}(i,i)}} - \frac{\mathbf{f}_{(m)}^j}{\sqrt{\mathbf{D}_{(m)}(j,j)}}\right\|_2^2 z_{ji}. \tag{15}$$

Then Eq. (9) with respect to $\mathbf{Z}$ is equal to:

$$\min_{\mathbf{Z}^\top\mathbf{1}=\mathbf{1},\mathbf{Z}\geq\mathbf{0}} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{p=1}^{v}\beta_p^2\left(z_{ji}^2 - 2c_{pij}z_{ji}\right) + \lambda t_{ij}z_{ji}, \tag{16}$$

where $\mathbf{C}_p = \mathbf{X}_p^\top\mathbf{W}_p\mathbf{A}$ and $t_{ij} = \left\|\frac{\mathbf{f}_{(n)}^i}{\sqrt{\mathbf{D}_{(n)}(i,i)}} - \frac{\mathbf{f}_{(m)}^j}{\sqrt{\mathbf{D}_{(m)}(j,j)}}\right\|_2^2$. Denoting $\mathbf{Z}_{:,i}$ as a vector with the i-th element to be $\mathbf{z}_{ji}$, we optimize $\mathbf{Z}$ by column:

$$\min_{\mathbf{Z}_{:,i}^\top\mathbf{1}=1,\mathbf{Z}_{:,i}\geq\mathbf{0}} \left\|\mathbf{Z}_{:,i}^\top - \left(\mathbf{C}_{i,:} - \frac{\lambda}{2}\mathbf{t}_{i,:}\right)/\sum_{p=1}^{v}\beta_p^2\right\|_2^2. \tag{17}$$

where $\mathbf{C} = \sum_{p=1}^{v}\beta_p^2\mathbf{C}_p$.

According to Nie, Wang, and Huang (2014), Eq. (17) has a closed form solution. We summarize the procedures of solving Eq. (7) in Algorithm 1.

Algorithm 1 Algorithm for optimizing $\mathbf{Z}$.

Input: $\mathbf{C} \in \mathbb{R}^{n\times m}$, cluster number k, a small value λ.
Output: $\mathbf{Z} \in \mathbb{R}^{m\times n}$ with exactly k connected components.

Initialize $\mathbf{F}_{(n)}$ and $\mathbf{F}_{(m)}$, which are formed by the k left and right singular vectors of $(\frac{1}{m}\mathbf{1}_n\mathbf{1}_m^\top)(\frac{n}{m}\mathbf{I}_m)^{-\frac{1}{2}}$ corresponding to the k largest singular values.
repeat
  For each i, update the i-th column of $\mathbf{Z}$ by solving Eq. (17), where the j-th element of $\mathbf{t}_{i,:}$ is $t_{ij} = \left\| \frac{\mathbf{f}^i_{(n)}}{\sqrt{\mathbf{D}_{(n)}(i,i)}} - \frac{\mathbf{f}^j_{(m)}}{\sqrt{\mathbf{D}_{(m)}(j,j)}} \right\|_2^2$;
  Update $\mathbf{D}_{(n)}$ and $\mathbf{D}_{(m)}$ where $d_{(n)}(i,i) = \sum_{j=1}^{m} z_{ji}$ and $d_{(m)}(j,j) = \sum_{i=1}^{n} z_{ji}$;
  Update $\mathbf{F}_{(n)}$ and $\mathbf{F}_{(m)}$ by solving Eq. (12);
  Update λ due to $\widetilde{\mathbf{L}}$;
until $\mathbf{Z}$ has exactly k connected components.

Optimize $\mathbf{W}_p$ by fixing other variables With $\mathbf{Z}, \mathbf{A}$ and $\boldsymbol{\beta}$ being fixed, the optimization for $\mathbf{W}_p$ can be formulated as

$$\min_{\mathbf{W}_p} \sum_{p=1}^{v} \beta_p^2 \left\|\mathbf{X}_p - \mathbf{W}_p\mathbf{A}\mathbf{Z}\right\|_F^2, s.t. \mathbf{W}_p^\top\mathbf{W}_p = \mathbf{I}_l. \quad (18)$$

Since $\mathbf{W}_p$ is independent in each view, Eq. (18) can be transformed into the following equivalent problem

$$\max_{\mathbf{W}_p} \mathrm{Tr}\left(\mathbf{W}_p^\top\mathbf{G}_p\right), s.t. \mathbf{W}_p^\top\mathbf{W}_p = \mathbf{I}_l, \quad (19)$$

where $\mathbf{G}_p = \mathbf{X}_p\mathbf{Z}^\top\mathbf{A}^\top$. The optimal solution of $\mathbf{W}_p$ is $\mathbf{U}\mathbf{V}^\top$ where $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular matrix of $\mathbf{G}_p$.

Optimize $\mathbf{A}$ by fixing other variables Fixing $\mathbf{Z}, \mathbf{W}_p$ and $\boldsymbol{\beta}$, the optimization of $\mathbf{A}$ can be rewritten as,

$$\min_{\mathbf{A}} \sum_{p=1}^{v} \beta_p^2 \left\|\mathbf{X}_p - \mathbf{W}_p\mathbf{A}\mathbf{Z}\right\|_F^2, s.t. \mathbf{A}^\top\mathbf{A} = \mathbf{I}_m. \quad (20)$$

Similar to the optimization process of $\mathbf{W}_p$, the above equation equals to the following form by removing unrelated items,

$$\max_{\mathbf{A}} \mathrm{Tr}\left(\mathbf{A}^\top\mathbf{B}\right), s.t. \mathbf{A}^\top\mathbf{A} = \mathbf{I}_m, \quad (21)$$

where $\mathbf{B} = \sum_{p=1}^{v} \beta_p^2\mathbf{W}_p^\top\mathbf{X}_p\mathbf{Z}^\top$. Supposing $\mathbf{U}\Sigma\mathbf{V}^\top$ to be the singular value decomposition result of $\mathbf{B}$, the optimal solution of Eq. (21) is $\mathbf{U}\mathbf{V}^\top$.

Optimize $\boldsymbol{\beta}$ by fixing other variables When $\mathbf{Z}, \mathbf{W}_p$ and $\mathbf{A}$ are fixed, the objective function with respect to $\boldsymbol{\beta}$ can be formulated as

$$\min_{\boldsymbol{\beta}} \sum_{p=1}^{v} \beta_p^2\mathbf{M}_p^2, s.t. \boldsymbol{\beta}^\top\mathbf{1} = 1, \boldsymbol{\beta} \geq 0, \quad (22)$$

where $\mathbf{M}_p = \left\|\mathbf{X}_p - \mathbf{W}_p\mathbf{A}\mathbf{Z}\right\|_F$. We can obtain the optimal β_p by Cauchy-Buniakowsky-Schwarz inequality as

$$\beta_p = \frac{\frac{1}{\mathbf{M}_p}}{\sum_{p=1}^{v} \frac{1}{\mathbf{M}_p}}, \quad (23)$$

The whole pipeline of solving Eq. (6) is summarized in Algorithm 2.

Algorithm 2 EOMSC-CA

Input: Multi-view dataset $\{\mathbf{X}_p\}_{p=1}^{v}$, cluster number k.
Output: $\mathbf{Z} \in \mathbb{R}^{m\times n}$ with exactly k connected components.

Initialize $\mathbf{W}$,$\mathbf{A}$,$\mathbf{Z}$. Initialize $\beta_p = \frac{1}{v}$.
repeat
  Compute $\mathbf{C} = \sum_{p=1}^{v} \beta_p^2\mathbf{X}_p^\top\mathbf{W}_p\mathbf{A}$.
  Update $\mathbf{Z}$ by Algorithm 1;
  Update $\mathbf{W}_p$ by solving Eq. (19);
  Update $\mathbf{A}$ by solving Eq. (21);
  Update $\boldsymbol{\beta}$ by solving Eq. (23);
until converged.

Complexity Analysis

Due to the application of anchor strategy, EOMSC-CA has a low computational complexity. Specifically, it takes $\mathcal{O}\left(nm^2t + m^3t + nmkt\right)$ to perform Algorithm 1 to construct $\mathbf{Z}$ with t being the iteration number, where solving Eq. (12) costs $\mathcal{O}\left(nm^2 + m^3\right)$ and solving Eq. (17) costs $\mathcal{O}\left(nmk\right)$ in each iteration. In the process of optimizing $\mathbf{W}_p$, performing SVD in each view needs $\mathcal{O}\left(d_pl^2\right)$ and matrix multiplication needs $\mathcal{O}\left(d_plk^2\right)$. When updating $\mathbf{A}$, it cost $\mathcal{O}\left(ml^2\right)$ for SVD and $\mathcal{O}\left(mlk^2\right)$ for matrix multiplication. And calculating β_p costs only $\mathcal{O}\left(1\right)$. Totally, the main computational complexity in Algorithm 2 is $\mathcal{O}\left(dk^2 + dk^3 + nk^2t\right)$. In our algorithm, $d \ll n$, $k \ll n$ and $t \ll n$. Therefore, the optimization step in our algorithm is a linear complexity $\mathcal{O}\left(n\right)$.

After obtaining a k-connected graph $\mathbf{Z}$, we perform a linear graph algorithm on it instead of performing SVD on graph and then k-means to get results. The computational complexity of this step is $\mathcal{O}\left(nmk\right)$, which is also a linear complexity respect to the number of samples. By contrast, most of the multi-view subspace clustering algorithms have $\mathcal{O}\left(n^3\right)$ complexity.

Table 2: Datasets used in our experiments.

Dataset	#Samples	#View	#Class
ORL_mtv	400	3	40
Caltech101-7	1474	6	7
Mfeat	2000	6	10
Caltech101-20	2386	6	20
Caltech101-all	9144	5	102
SUNRGBD	10335	2	45
NUSWIDEOBJ	30000	5	31
AwA	30475	6	50
YoutubeFace	101499	5	31

Experiments

In the section, we conduct experiments to evaluate the performance of the proposed method on multi-view datasets.

Table 3: ACC, NMI and Fscore comparison of different clustering algorithms on datasets.

Dataset	MLRSSC (2018)	AMGL (2016)	SFMC (2020)	RMKM (2013)	BMVC (2018)	LMVSC (2020)	MSGL (2021)	Ours
Hyper-parameters number	3	1	0	1	4	1	2	0
ACC								
ORL_mtv	0.0500	**0.6951**	0.6150	0.4350	0.4875	0.5862	0.2100	**0.6225**
Caltech101-7	0.6090	0.3960	0.6526	0.2877	0.2239	0.3535	**0.7347**	**0.8351**
Mfeat	0.2000	**0.8212**	0.7575	0.6710	0.6935	0.8170	0.7545	**0.8220**
Caltech101-20	0.3600	0.2878	**0.5947**	0.3961	0.1689	0.2929	0.4790	**0.6404**
Caltech101-all	0.1085	0.1401	0.1777	0.1650	**0.2123**	0.1449	0.1412	**0.2219**
SUNRGBD	0.1391	0.0981	0.1136	0.1771	0.1669	**0.1809**	0.1310	**0.2322**
NUSWIDEOBJ	N/A	N/A	0.1221	0.1328	0.1299	**0.1476**	0.1204	**0.1968**
AWA	N/A	N/A	0.0390	0.0656	**0.0867**	0.0723	0.0802	**0.0870**
YotubeFace	N/A	N/A	N/A	N/A	0.0897	0.1403	**0.1671**	**0.2664**
NMI								
ORL_mtv	0.1583	**0.8717**	**0.8269**	0.7010	0.6773	0.7880	0.4372	**0.8815**
Caltech101-7	0.1788	0.4408	**0.5629**	0.1411	0.0470	0.3384	0.3794	**0.5219**
Mfeat	0.2863	**0.8674**	**0.8684**	0.6533	0.6605	0.7609	0.7654	**0.8319**
Caltech101-20	0.2008	0.4760	0.4285	**0.5034**	0.1626	0.4187	0.3113	**0.5109**
Caltech101-all	0.0474	**0.3529**	0.2613	0.3494	**0.4246**	0.3332	0.2612	0.2533
SUNRGBD	0.0421	0.1840	0.0230	**0.2531**	0.1954	**0.2550**	0.0933	**0.2333**
NUSWIDEOBJ	N/A	N/A	0.0096	**0.1435**	0.1290	0.1276	0.0573	**0.1327**
AWA	N/A	N/A	0.0034	0.0738	**0.1372**	0.0855	0.0792	**0.0972**
YotubeFace	N/A	N/A	N/A	N/A	**0.0593**	**0.1179**	0.0007	0.0022
Fscore								
ORL_mtv	0.0582	**0.5123**	0.3066	0.3068	0.3054	0.4599	0.0517	**0.6200**
Caltech101-7	0.5287	0.4035	0.6409	0.2879	0.2275	0.3790	**0.6524**	**0.7967**
Mfeat	0.2739	**0.8083**	0.7111	0.5922	0.5879	0.7252	0.7011	**0.7701**
Caltech101-20	0.3069	0.2182	0.3150	0.3565	0.1138	0.2564	**0.4174**	**0.6471**
Caltech101-all	0.0502	0.0406	0.0462	**0.1486**	**0.1854**	0.1047	0.0864	0.1287
SUNRGBD	**0.1291**	0.0644	0.1212	0.1168	0.1019	0.1159	0.0949	**0.1507**
NUSWIDEOBJ	N/A	N/A	**0.1140**	0.0865	0.0881	0.0932	0.0856	**0.1367**
AWA	N/A	N/A	0.0457	0.0359	**0.0559**	0.0365	0.0421	**0.0599**
YotubeFace	N/A	N/A	N/A	N/A	0.0579	0.0831	**0.1511**	**0.1642**

Benchmark Datasets

We perform experiments on nine widely used multi-view benchmark datasets: ORL_mtv, Caltech101-7, Mfeat, Caltech101-20, Caltech101-all, SUNRGBD, NUSWIDEOBJ, AwA, YoutubeFace. The details of them are shown in Table 2. Specifically, ORL_mtv contains 400 images in 40 classes. Caltech101-7 with 1474 instances in 7 categories and Caltech101-20 with 2386 subjects in 20 classes are both subsets of the image dataset Caltech101 (Fei-Fei, Fergus, and Perona 2004). Mfeat was generated from UCI machine learning repository which consists of the digits from 0 to 9. SUNRGBD (Song, Lichtenberg, and Xiao 2015) consists of 10335 indoor scene images spread over 45 classes. NUSWIDEOBJ (Chua et al. 2009) is a object recognition database with 30000 objects. AwA contains 50 different animals with their six features. YoutubeFace is produced from YouTube with 101499 instances.

Experimental Setup

We compare our method with the following seven state-of-the-art multi-view clustering methods: **Multi-view Low-rank Sparse Subspace Clustering (MLRSSC)** (Brbić and Kopriva 2018); **Parameter-Free Auto-Weighted Multiple Graph Learning: A Framework for Multiview Clustering and Semi-Supervised Classification (AMGL)** (Nie et al. 2016a); **Multi-view Clustering: A Scalable and Parameter-free Bipartite Graph Fusion Method (SFMC)** (Li et al. 2020); **Multi-View K-Means Clustering on Big Data (RMKM)** (Cai, Nie, and Huang 2013); **Binary Multi-View Clustering (BMVC)** (Zhang et al. 2018); **Large-scale Multi-view Subspace Clustering in Linear Time (LMVSC)** (Kang et al. 2020); **Structured Graph Learning for Scalable Subspace Clustering: From Single View to Multiview (MSGL)** (Kang et al. 2021).

The proposed method has no hyper-parameters to be tuned, but we need to determine the anchors number and the dimension of anchor matrix. In the experiments, the anchors number and the dimension of anchor matrix are both traverse $[k, 2k, ..., 7k]$ where k is the clustering number of each dataset. For the compared algorithms, we search their best parameters for fairness. Moreover, we run 50 times k-means and report the best result. To evaluate the clustering performance, we employ three widely used criteria including accuracy (ACC), normalized mutual information (NMI) and Fscore. All the experiments are performed on a desktop with Intel core i9-10900X CPU and 64G RAM, MATLAB 2019b(64-bit).

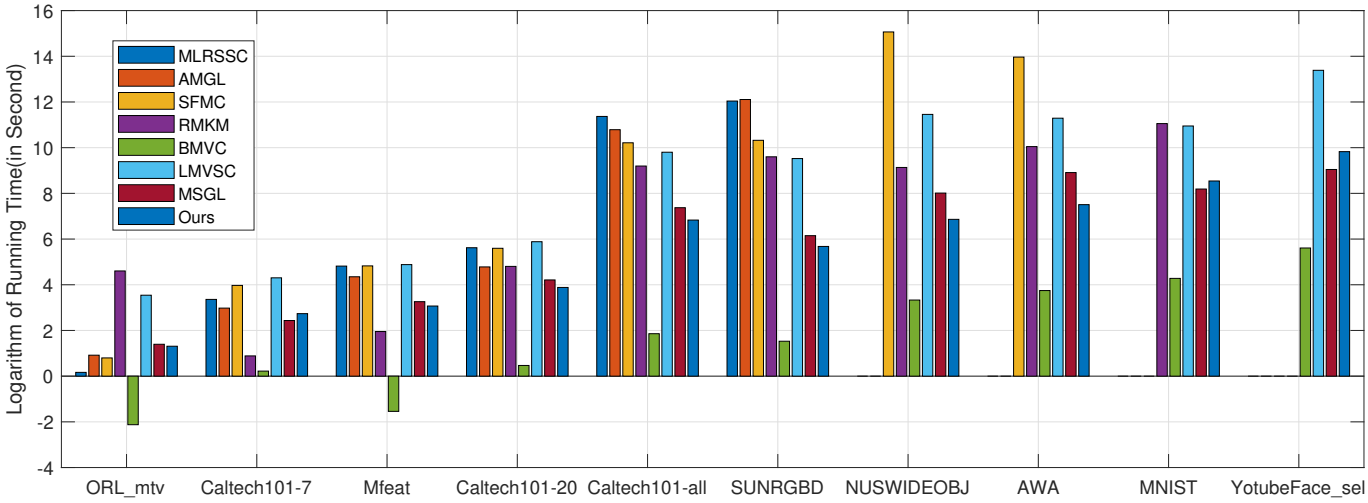
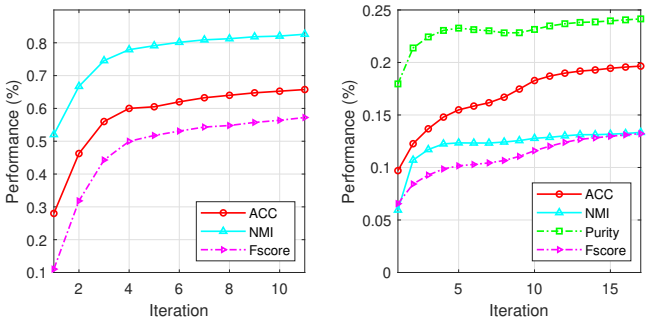
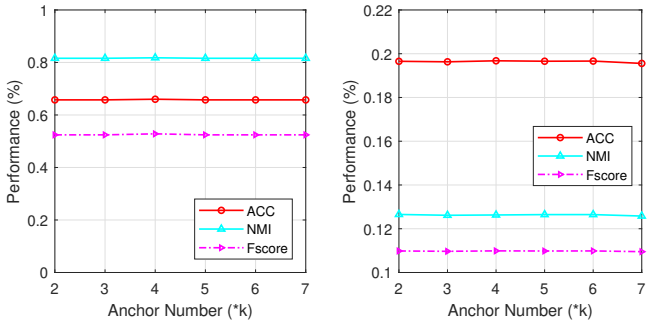


Figure 2: The running time of the compared algorithms on the benchmark datasets respectively.



(a) ORL_mtv (b) NUSWIDEOBJ

Figure 3: The clustering ACC, NMI and Fscore of our algorithm w.r.t the number of iterations.



(a) ORL_mtv (b) NUSWIDEOBJ

Figure 4: Sensitivity analysis of anchor number for our method over ORL_mtv and NUSWIDEOBJ.

Experimental Results

Table 3 compares the clustering performance of the EOMSC-CA with other methods on nine benchmark datasets. We mark the optimal results in red, use bold for the sub-optimal and close results, and 'N/A' to indicate out-of-memory issue. According to the results, we have the following conclusions:

- The proposed algorithm can achieve the best on the eight datasets and is close to the best on the rest datasets, which proves its effectiveness in multi-view clustering.
- AMGL has the best performance among the compared methods on ORL_mtv. However, AMGL cannot be applied to large-scale datasets due to high complexity. Compared with the above method, EOMSC-CA not only has close performance on the corresponding datasets, but also can be performed on large-scale datasets.

The comparison of NMI and Fscore are also reported in Table 3. We observe that the proposed algorithm has promising performance among datasets. To further prove the effectiveness of our algorithm optimization, we also conduct experiments on evolution of the clustering performance with variation iterations. As shown in Figure 3, the ACC, NMI and Fscore monotonically increase at each iteration and stabilize in the last few iterations. The results verify EOMSC-CA's effectiveness in optimization.

Running Time Comparison

The running time comparisons of various algorithms on nine multi-view datasets are recorded in Figure 2. Compared with most state-of-the-art multi-view methods, EOMSC-CA has the better computational efficiency as shown in the figure. Although BMVC and MSGL have less running time, they have four and two hyper-parameters to be choose with poorer clustering performance. Therefore, it can be demonstrated that our method has high efficiency on multi-view clustering, which is proved by theoretical and experimental results.

Sensitivity analysis

In order to analyze the influence of the number of anchor points on the efficiency of the algorithm, We fixed l and conducted a comparative experiment on ORL_mtv and NUSWIDEOBJ. As shown in Figure 4, our algorithm is not greatly affected by the anchor number.

Conclusion

In this paper, we propose a novel multi-view subspace clustering termed as EOMSC-CA. Different from previous anchor-based methods, the selection of anchor points and the construction of subspace graphs are optimized jointly to improve the clustering performance in our method. Moreover, by imposing a connectivity constraint, the constructed graph can directly get the final clustering result without post-processing. Extensive experiments on real-world datasets prove the effectiveness and efficiency of the proposed method.

Acknowledgements

This work was supported by the National Key R&D Program of China (project no. 2020AAA0107100) and the National Natural Science Foundation of China (project no. 61922088, 61906020, 61872371 and 62006237).

References

Adler, A.; Elad, M.; and Hel-Or, Y. 2015. Linear-time subspace clustering via bipartite graph modeling. *IEEE transactions on neural networks and learning systems* 26(10): 2234–2246.

Brbić, M.; and Kopriva, I. 2018. Multi-view low-rank sparse subspace clustering. *Pattern Recognition* 73: 247–258.

Cai, X.; Nie, F.; and Huang, H. 2013. Multi-view k-means clustering on big data. In *Twenty-Third International Joint conference on artificial intelligence*.

Cao, X.; Zhang, C.; Fu, H.; Liu, S.; and Zhang, H. 2015. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–594.

Chen, X.; and Cai, D. 2011. Large scale spectral clustering with landmark-based representation. In *Twenty-fifth AAAI conference on artificial intelligence*.

Chua, T.-S.; Tang, J.; Hong, R.; Li, H.; Luo, Z.; and Zheng, Y. 2009. Nus-wide: a real-world web image database from national university of singapore. In *Proceedings of the ACM international conference on image and video retrieval*, 1–9.

Ding, Z.; and Fu, Y. 2014. Low-rank common subspace for multi-view learning. In *2014 IEEE international conference on Data Mining*, 110–119. IEEE.

Elhamifar, E.; and Vidal, R. 2013. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE transactions on pattern analysis and machine intelligence* 35(11): 2765–2781.

Fan, K. 1949. On a theorem of Weyl concerning eigenvalues of linear transformations I. *Proceedings of the National Academy of Sciences of the United States of America* 35(11): 652.

Fei-Fei, L.; Fergus, R.; and Perona, P. 2004. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *2004 conference on computer vision and pattern recognition workshop*, 178–178. IEEE.

Gao, H.; Nie, F.; Li, X.; and Huang, H. 2015. Multi-view subspace clustering. In *Proceedings of the IEEE international conference on computer vision*, 4238–4246.

Gao, Q.; Xia, W.; Wan, Z.; Xie, D.; and Zhang, P. 2020. Tensor-SVD based graph learning for multi-view subspace clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3930–3937.

Kang, Z.; Guo, Z.; Huang, S.; Wang, S.; Chen, W.; Su, Y.; and Xu, Z. 2019. Multiple Partitions Aligned Clustering. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 2701–2707. International Joint Conferences on Artificial Intelligence Organization.

Kang, Z.; Lin, Z.; Zhu, X.; and Xu, W. 2021. Structured Graph Learning for Scalable Subspace Clustering: From Single View to Multiview. *IEEE Transactions on Cybernetics* .

Kang, Z.; Zhou, W.; Zhao, Z.; Shao, J.; Han, M.; and Xu, Z. 2020. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 4412–4419.

Li, X.; Zhang, H.; Wang, R.; and Nie, F. 2020. Multi-view clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* .

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2012. Robust recovery of subspace structures by low-rank representation. *IEEE transactions on pattern analysis and machine intelligence* 35(1): 171–184.

Liu, J.; Liu, X.; Yang, Y.; Guo, X.; Kloft, M.; and He, L. 2021a. Multiview subspace clustering via co-training robust data representation. *IEEE Transactions on Neural Networks and Learning Systems* .

Liu, X.; Liu, L.; Liao, Q.; Wang, S.; Zhang, Y.; Tu, W.; Tang, C.; Liu, J.; and Zhu, E. 2021b. One Pass Late Fusion Multi-view Clustering. In *International Conference on Machine Learning*, 6850–6859. PMLR.

Liu, X.; Zhou, S.; Liu, L.; Tang, C.; Wang, S.; Liu, J.; and Zhang, Y. 2021c. Localized Simple Multiple Kernel K-Means. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9293–9301.

Liu, X.; Zhu, X.; Li, M.; Tang, C.; Zhu, E.; Yin, J.; and Gao, W. 2019. Efficient and effective incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 4392–4399.

Lu, C.; Yan, S.; and Lin, Z. 2016. Convex sparse spectral clustering: Single-view to multi-view. *IEEE Transactions on Image Processing* 25(6): 2833–2843.

Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, 849–856.

Nie, F.; Li, J.; Li, X.; et al. 2016a. Parameter-free auto-weighted multiple graph learning: a framework for multi-view clustering and semi-supervised classification. In *IJCAI*, 1881–1887.

Nie, F.; Wang, X.; Deng, C.; and Huang, H. 2017. Learning a structured optimal bipartite graph for co-clustering. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4132–4141.

Nie, F.; Wang, X.; and Huang, H. 2014. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 977–986.

Nie, F.; Wang, X.; Jordan, M.; and Huang, H. 2016b. The constrained laplacian rank algorithm for graph-based clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

Ou, Q.; Wang, S.; Zhou, S.; Li, M.; Guo, X.; and Zhu, E. 2020. Anchor-based multiview subspace clustering with diversity regularization. *IEEE MultiMedia* 27(4): 91–101.

Song, S.; Lichtenberg, S. P.; and Xiao, J. 2015. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 567–576.

Sun, M.; Zhang, P.; Wang, S.; Zhou, S.; Tu, W.; Liu, X.; Zhu, E.; and Wang, C. 2021. Scalable Multi-view Subspace Clustering with Unified Anchors. In *Proceedings of the 29th ACM International Conference on Multimedia*, 3528–3536.

Tang, C.; Zhu, X.; Liu, X.; Li, M.; Wang, P.; Zhang, C.; and Wang, L. 2018. Learning a joint affinity graph for multiview subspace clustering. *IEEE Transactions on Multimedia* 21(7): 1724–1736.

Vidal, R. 2011. Subspace clustering. *IEEE Signal Processing Magazine* 28(2): 52–68.

Wang, C.-D.; Chen, M.-S.; Huang, L.; Lai, J.-H.; and Philip, S. Y. 2020. Smoothness regularized multiview subspace clustering with kernel learning. *IEEE Transactions on Neural Networks and Learning Systems* .

Wang, S.; Liu, X.; Liu, L.; Zhou, S.; and Zhu, E. 2021. Late Fusion Multiple Kernel Clustering With Proxy Graph Refinement. *IEEE Transactions on Neural Networks and Learning Systems* .

Wang, S.; Liu, X.; Zhu, E.; Tang, C.; Liu, J.; Hu, J.; Xia, J.; and Yin, J. 2019. Multi-view Clustering via Late Fusion Alignment Maximization. In *IJCAI*, 3778–3784.

Wang, Y.; Wu, L.; Lin, X.; and Gao, J. 2018. Multiview spectral clustering via structured low-rank matrix factorization. *IEEE transactions on neural networks and learning systems* 29(10): 4833–4843.

Xu, N.; Guo, Y.; Zheng, X.; Wang, Q.; and Luo, X. 2018. Partial multi-view subspace clustering. In *Proceedings of the 26th ACM International conference on multimedia*, 1794–1801.

Zhan, K.; Zhang, C.; Guan, J.; and Wang, J. 2017. Graph learning for multiview clustering. *IEEE transactions on cybernetics* 48(10): 2887–2895.

Zhang, C.; Fu, H.; Liu, S.; Liu, G.; and Cao, X. 2015. Low-rank tensor constrained multiview subspace clustering. In *Proceedings of the IEEE international conference on computer vision*, 1582–1590.

Zhang, G.-Y.; Zhou, Y.-R.; Wang, C.-D.; Huang, D.; and He, X.-Y. 2021. Joint representation learning for multi-view subspace clustering. *Expert Systems with Applications* 166: 113913.

Zhang, P.; Liu, X.; Xiong, J.; Zhou, S.; Zhao, W.; Zhu, E.; and Cai, Z. 2020. Consensus One-step Multi-view Subspace Clustering. *IEEE Transactions on Knowledge and Data Engineering* .

Zhang, Z.; Liu, L.; Shen, F.; Shen, H. T.; and Shao, L. 2018. Binary multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence* 41(7): 1774–1782.

Zhou, S.; Liu, X.; Li, M.; Zhu, E.; Liu, L.; Zhang, C.; and Yin, J. 2020a. Multiple Kernel Clustering With Neighbor-Kernel Subspace Segmentation. *IEEE Transactions on Neural Networks and Learning Systems* 31(4): 1351–1362. doi: 10.1109/TNNLS.2019.2919900.

Zhou, S.; Liu, X.; Liu, J.; Guo, X.; Zhao, Y.; Zhu, E.; Zhai, Y.; Yin, J.; and Gao, W. 2020b. Multi-view spectral clustering with optimal neighborhood Laplacian matrix. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 6965–6972.

Zhou, S.; Ou, Q.; Liu, X.; Wang, S.; Liu, L.; Wang, S.; Zhu, E.; Yin, J.; and Xu, X. 2021. Multiple Kernel Clustering With Compressed Subspace Alignment. *IEEE Transactions on Neural Networks and Learning Systems* 1–12. doi:10.1109/TNNLS.2021.3093426.