

# An Effective Augmented Lagrangian Method for Fine-Grained Multi-View Optimization

Yuze Tan<sup>1,2</sup>, Hecheng Cai<sup>1,2</sup>, Shudong Huang<sup>1,2\*</sup>, Shuping Wei<sup>3</sup>, Fan Yang<sup>1,4</sup>, Jiancheng Lv<sup>1,2</sup>

<sup>1</sup>College of Computer Science, Sichuan University, Chengdu 610065, China

<sup>2</sup>Engineering Research Center of Machine Learning and Industry Intelligence, Ministry of Education, Chengdu, 610065, China

<sup>3</sup>Nuclear Power Institute of China

<sup>4</sup>Sichuan IotDT Technology Co., Ltd.

{tanyuze, caihecheng}@stu.scu.edu.cn, {huangsd, lvjiancheng}@scu.edu.cn, weishuping@npic.ac.cn, yang.fan@live.com

## Abstract

The significance of multi-view learning in effectively mitigating the intricate intricacies entrenched within heterogeneous data has garnered substantial attention in recent years. Notwithstanding the favorable achievements showcased by recent strides in this area, a confluence of noteworthy challenges endures. To be specific, a majority of extant methodologies unceremoniously assign weights to data points view-wisely. This ineluctably disregards the intrinsic reality that disparate views confer diverse contributions to each individual sample, consequently neglecting the rich wellspring of sample-level structural insights harbored within the dataset. In this paper, we proposed an effective Augmented Lagrangian MethOd for fiNe-grained (ALMOND) multi-view optimization. This innovative approach scrutinizes the interplay among multiple views at the granularity of individual samples, thereby fostering the enhanced preservation of local structural coherence. The Augmented Lagrangian Method (ALM) is elaborately incorporated into our framework, which enables us to achieve an optimal solution without involving an inexplicable intermediate variable as previous methods do. Empirical experiments on multi-view clustering tasks across heterogeneous datasets serve to incontrovertibly showcase the effectiveness of our proposed methodology, corroborating its preeminence over incumbent state-of-the-art alternatives.

## Introduction

The proliferation of multi-view data sources has accentuated the need for methodologies that can distill valuable insights from a multitude of vantage points. For example, autonomous driving requires both RGB videos and corresponding dense 3D point clouds to construct more accurate real-time traffic information (Huang et al. 2018). Disease diagnosis takes health history data, physical exams as well as blood tests into consideration (Wang et al. 2020). Meanwhile, the image dataset can be described by multiple distinctive descriptors (Gao et al. 2015), such as LBP (Ojala, Pietikainen, and Maenpaa 2002), SIFT (Lowe

2004), and HOG (Dalal and Triggs 2005). However, traditional single-view clustering methods, proficient at segregating data within isolated modalities, frequently fall short in encapsulating the comprehensive knowledge embedded across multiple perspectives.

Among all existing multi-view clustering (MVC) techniques, graph-based methods have drawn the most attention due to their effectiveness and ease of implementation. Normally, graph-based MVC approaches first learn a similarity graph  $\mathbf{S}^{(v)}$  for each view, and then deploy different fusion strategies on these view-oriented graphs to obtain an optimal consensus graph  $\mathbf{S}$ , which can be further processed by two mainstream techniques. One is to restrict the rank of  $\mathbf{S}$  to be  $n - k$ , where  $n$  is the number of data points, while  $k$  is the cluster number. Since there are now exactly  $k$  connected components respecting  $k$  clusters in the optimal consensus graph  $\mathbf{S}$ , no post-processing is needed to acquire the clustering result. The other one is followed by conducting spectral clustering or  $k$ -means on the obtained consensus graph  $\mathbf{S}$  to produce the final result.

By taking advantage of the diffusion process, (Tang et al. 2020) proposed a parameter-free MVC method via cross-view graph diffusion. Since the original graph could be noisy or incomplete and is not directly applicable, (Pan and Zhao 2021) proposed a multi-view contrastive graph clustering method to learn a consensus graph. In (Huang et al. 2021, 2022b), they proposed to leverage the multi-view consistency and the multi-view diversity in a unified framework yielding a pure graph for each view. Multi-view anchor graph clustering could be extremely difficult since anchors are not consistent in feature dimensions. To this end, (Siwei et al. 2022) proposed the generalized and flexible anchor graph fusion framework. To utilize the multi-view information, (Wang, Pei, and Zhan 2022) designed a specific graph learning method by introducing graph regularization and local structure fusion patterns. In order to simultaneously consider the similarity of inter-view and intra-view, (Xia et al. 2023) proposed a variance-based de-correlation anchor selection strategy for bipartite construction. To extract high-level view-common information and reduce this influence of non-homophilous edges, (Ling et al. 2023) proposed dual label-guided graph refinement for multi-view graph cluster-

\*Corresponding author.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ing. (Liu et al. 2023) proposed a novel kernel function based on the emerging contrastive learning framework to capture complementary information across views better.

Although promising results have been demonstrated in above mentioned works, it is noteworthy that existing graph-based MVCs mainly focus on obtaining the consensus graph by fusing multiple view-oriented similarity graphs in a view-level manner. Hardly of them have considered the intersections of the heterogeneous information from a fine-grained view, let alone construct a self-consistent model, which shouldn't introduce any inexplicable intermediate variable. In this paper, we proposed an effective Augmented Lagrangian Method for fine-grained (ALMOND) multi-view optimization. Instead of integrating diverse information view-wisely, which would lead to an overlap of redundancies, resulting in a less precise shared cluster structure, we decided to go more fine-grained. Given that the alignment between multiple views is evident in the generally congruent local structures, with the divergent patterns being a minority, we delve deeper into the intersections of these perspectives at the sample level. This approach enhances the preservation of cross-view consistency. Furthermore, an effective augmented lagrangian method was introduced into our model so that no inexplicable intermediate variable exists, which preserves the self-consistency of our model. Note that the fine-grained fusion framework we proposed in this paper consists of several features that are distinct from existing methods.

- Rather than obtaining the consensus graph by fusing multiple similarity graphs view-wisely, we developed a fine-grained fusion strategy, which considered the fusion process at the sample level.
- Our method is self-consistent, which means an effective Augmented Lagrangian Method was introduced into our model such that it can achieve an optimal solution without involving an inexplicable intermediate variable as previous methods do. Therefore, the self-consistency can be better maintained.
- Extensive experiments on several benchmark datasets and corresponding results have proved the superiority of our proposed method.

## Related Work

### Multi-view Subspace Clustering

Multi-view subspace clustering has attracted widespread attention in recent years. It is based on the assumption that each data point can be expressed as a linear combination of other data points, which reduces the dimension of feature space. For multi-view data  $\{\mathbf{X}^{(v)}\}_{v=1}^m$ , where  $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$  denotes the  $v$ -th view data with  $d_v$  dimension. Therefore, a standard multi-view subspace clustering framework

Notation	Definition
$n$	Amount of samples
$m$	Amount of views
$d_v$	Dimension of $v$ -th view
$\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$	Data matrix
$\mathbf{Z}^{(v)} \in \mathbb{R}^{n \times n}$	Subspace representation
$\tilde{\mathbf{Z}} \in \mathbb{R}^{m \times n \times n}$	Concatenation of all $\mathbf{Z}^{(v)}$
$\tilde{\mathbf{Z}}_i \in \mathbb{R}^{m \times n}$	$i$ -th frontal slice of $\tilde{\mathbf{Z}}$
$\alpha, \beta, \gamma$	Balance parameters

Table 1: Mainly used notations

can be expressed as follow:

$$\begin{aligned} \min_{\mathbf{Z}^{(v)}, \mathbf{S}} \sum_{v=1}^m \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \alpha g \left( \mathbf{Z}^{(v)} \right) \\ + \phi f \left( \mathbf{S}, \mathbf{Z}^{(v)} \right) \\ \text{s.t. } \mathbf{Z}^{(v)} \geq 0, \text{diag} \left( \mathbf{Z}^{(v)} \right) = 0, \end{aligned} \quad (1)$$

where  $\alpha$  and  $\phi$  are balance parameters,  $\mathbf{Z}^{(v)}$  is the learned subspace corresponding to each view, while  $g(\cdot)$  and  $f(\cdot)$  represent the penalty term and customized graph fusion strategy respectively. Hence final results can be reached by performing spectral clustering on the consensus graph  $\mathbf{S}$ .

Based on the multi-view subspace clustering framework, a plethora of works that have promising results were proposed. (Wang et al. 2021) proposed a novel subspace clustering method that jointly conducted anchor selection and subspace graph construction into a unified optimization formulation. To address the problem that most anchor-based multi-view subspace clustering methods adopt fixed anchor points separating from the subsequential anchor graph construction, (Liu et al. 2022) combined anchor learning and graph construction into a uniform framework to boost clustering performance. Meanwhile, (Tan et al. 2023b) proposed to integrate metric learning and graph learning for multi-view clustering so that noise and diverse relationships within data can be tackled, to list a few.

### Graph Fusion Strategy

How to combine diverse and rich information from multiple views has played a significant role in multi-view clustering. (Gao et al. 2015) conducted multi-view graph fusion by imposing consensus graph  $\mathbf{S}$  to be an approximation of learned subspaces  $\mathbf{Z}^{(v)}$  as follows:

$$\begin{aligned} \min_{\mathbf{Z}^{(v)}} \sum_{v=1}^m \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \alpha \left\| \mathbf{Z} - \mathbf{Z}^{(v)} \right\|_F^2 \\ \text{s.t. } \mathbf{Z}^{(v)} \geq 0, \text{diag} \left( \mathbf{Z}^{(v)} \right) = 0. \end{aligned} \quad (2)$$

It is obvious that the above approach equally treats each individual subspace and ignores the different contributions of different views. Thus, (Nie, Li, and Li 2016; Nie et al. 2017)

proposed the auto-weighted framework for multiple graph learning, which can be concluded as:

$$\begin{aligned} \min_{\mathbf{Z}^{(v)}} \sum_{v=1}^m \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \alpha^{(v)} \left\| \mathbf{Z} - \mathbf{Z}^{(v)} \right\|_F^2 \\ \text{s.t. } \mathbf{Z}^{(v)} \geq 0, \text{diag}(\mathbf{Z}^{(v)}) = 0, \end{aligned} \quad (3)$$

where

$$\alpha^{(v)} = \frac{1}{2 \left\| \mathbf{Z} - \mathbf{Z}^{(v)} \right\|}. \quad (4)$$

Aiming at achieving the learning of each view graph matrix and the learning of the unified graph matrix in a mutual reinforcement manner, (Wang, Yang, and Liu 2019) proposed graph-based multi-view clustering as the following form:

$$\begin{aligned} \min_{\{\mathbf{S}^v\}, \mathbf{U}} \sum_{v=1}^m \sum_{i,j=1}^n \left\| \mathbf{x}_i^v - \mathbf{x}_j^v \right\|_2^2 s_{ij}^v + \beta \sum_{v=1}^m \sum_i \left\| \mathbf{s}_i^v \right\|_2^2 \\ + \sum_{v=1}^m w_v \left\| \mathbf{U} - \mathbf{S}^v \right\|_F^2 \\ \text{s.t. } \forall v, s_{ii}^v = 0, s_{ij}^v \geq 0, \mathbf{1}^T \mathbf{s}_i^v = 1, \\ u_{ij} \geq 0, \mathbf{1}^T \mathbf{u}_i = 1, \text{rank}(\mathbf{L}_U) = n - c. \end{aligned} \quad (5)$$

This approach seems to take sample-wise information into account. Nevertheless, it still signs weights to multiple graphs in a view-wise manner.

A great number of multi-view graph-base clustering methods based on these fusion strategies including a variety of their variants, have been proposed afterward. (Tan et al. 2023a) integrated topological manifold learning with sample-level graph fusion, which effectively exploited the local structure of data, but still didn't overcome the problem of the inexplicable variable. (Yu et al. 2023) proposed sample-level weights learning for multi-view clustering on spectral rotations, which is essentially a two-stage method. (Wang, Pei, and Zhan 2022) proposed to consider the sample-wise fusion strategy. Yet, it failed to build a self-consistent model, which has introduced an inexplicable intermediate variable to its object function.

Therefore, none of the existing work has contributed a self-consistent fine-grained fusion strategy that integrates sample-wise graph fusion and multi-view subspace clustering seamlessly.

## The Proposed Method

An intuitive idea of implementing ALMOND graph fusion is signing weight to every instance in each view. As shown in Figure.1, in traditional fusion strategy, the consensus graph can be regarded as the linear combination of multiple latent representations with respect to different views. On the contrary, our fine-grained fusion strategy emphasizes the consistency of heterogeneous information mainly lay in the generally same local structure. Therefore, the weight of each sample in every view is given to calculate a more accurate consensus graph sample-wisely so that the inconsistency in-

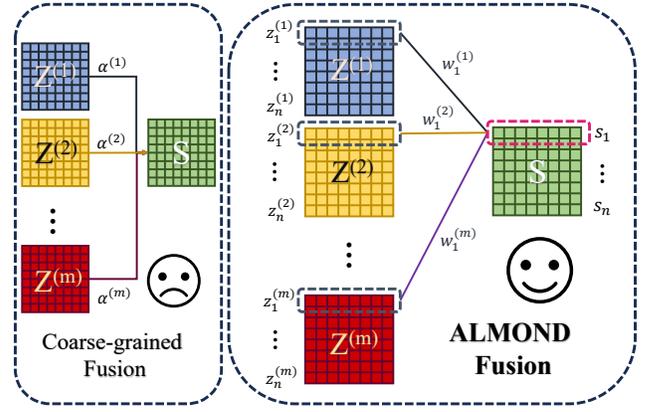


Figure 1: Our proposed ALMOND fusion strategy versus traditional coarse-grained fusion pattern.

produced by view-wise fusion can be suppressed. Consequently, the fine-grained fusion strategy can be modeled as:

$$\begin{aligned} \min_{\mathbf{S}} \sum_{i=1}^n \frac{\beta}{2} \left\| \mathbf{s}_i^T - \mathbf{w}_i^T \tilde{\mathbf{Z}}_i \right\|_2^2 + \frac{\gamma}{2} \left\| \mathbf{S} \right\|_F^2 \\ \text{s.t. } \mathbf{S} \geq 0, \mathbf{S} \mathbf{1} = \mathbf{1}, \mathbf{w}_i^T \mathbf{1} = 1, \end{aligned} \quad (6)$$

where  $\mathbf{s}_i$  is the  $i$ -th row of  $\mathbf{S}$ ,  $\mathbf{w}_i$  is the weight vector corresponding to the  $i$ -th sample. Meanwhile,  $\tilde{\mathbf{Z}} \in \mathbb{R}^{m \times n \times n}$  is a tensor which is concatenated by all  $\{\mathbf{Z}^{(v)}\}_{v=1}^m$  and  $\tilde{\mathbf{Z}}_i \in \mathbb{R}^{m \times n}$  is the  $i$ -th frontal slice of  $\tilde{\mathbf{Z}}$ . Thereby, the consensus graph  $\mathbf{S}$  that can depict more underlying structure of data is now obtained. Furthermore, with the help of multi-view subspace learning in Eq. (1), our proposed model can finally be formulated as:

$$\begin{aligned} \min_{\mathbf{Z}^{(v)}, \mathbf{J}^{(v)}, \mathbf{W}, \mathbf{S}, \mathbf{Y}^{(v)}} \sum_{v=1}^m \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \frac{\alpha}{2} \left\| \mathbf{Z}^{(v)} \right\|_F^2 \\ + \frac{\beta}{2} \sum_{i=1}^n \left\| \mathbf{s}_i^T - \mathbf{w}_i^T \tilde{\mathbf{Z}}_i \right\|_2^2 + \frac{\gamma}{2} \left\| \mathbf{S} \right\|_F^2 \\ \text{s.t. } \mathbf{Z}^{(v)} \geq 0, \mathbf{S} \geq 0, \mathbf{w}_i \geq 0, \mathbf{S} \mathbf{1} = \mathbf{1}, \mathbf{w}_i^T \mathbf{1} = 1. \end{aligned} \quad (7)$$

Note that  $\mathbf{Z}^{(v)}$  and  $\tilde{\mathbf{Z}}$  have intersections on a certain sample in a particular view which makes them highly coupled. Hence, it is impossible to optimize  $\mathbf{Z}^{(v)}$  nor  $\tilde{\mathbf{Z}}$  by adapting traditional alternative optimization procedures. In the next section, we propose to develop an effective Augmented Lagrangian MethOd for the fiNe-graineD (ALMOND) multi-view optimization, which delicately passes around the issue of inconsistency inherently and overcomes the difficulty of optimizing these two variables.

## Optimization

It's easy to see that  $\tilde{\mathbf{Z}}_i$  is dependent on all  $\{\mathbf{Z}^{(v)}\}_{v=1}^m$  and  $\mathbf{Z}^{(v)}$  in the first term cannot be broken down to sample form due to its multiplication with  $\mathbf{X}^{(v)}$ . Inspired by Augmented Lagrangian Method (ALM) (Andreani et al. 2008), we first

convert it to the following equivalent problem by introducing an auxiliary variable  $\mathbf{J}^{(v)}$  for each view:

$$\begin{aligned} & \min_{\mathbf{Z}^{(v)}, \mathbf{J}^{(v)}, \mathbf{W}, \mathbf{S}} \sum_{v=1}^m \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \\ & \frac{\alpha}{2} \left\| \mathbf{Z}^{(v)} \right\|_F^2 + \frac{\beta}{2} \sum_{i=1}^n \left\| \mathbf{s}_i^T - \mathbf{w}_i^T \tilde{\mathbf{J}}_i \right\|_2^2 + \frac{\gamma}{2} \left\| \mathbf{S} \right\|_F^2 \quad (8) \\ & \text{s.t. } \mathbf{Z}^{(v)} \geq 0, \mathbf{Z}^{(v)} = \mathbf{J}^{(v)} \\ & \mathbf{S} \geq 0, \mathbf{w}_i \geq 0, \mathbf{S} \mathbf{1} = \mathbf{1}, \mathbf{w}_i^T \mathbf{1} = \mathbf{1}, \end{aligned}$$

Hence, we can deploy ALM to sort the above problem. The corresponding augmented lagrangian function is:

$$\begin{aligned} & \min_{\mathbf{Z}^{(v)}, \mathbf{J}^{(v)}, \mathbf{W}, \mathbf{S}, \mathbf{Y}^{(v)}} \sum_{v=1}^m \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \\ & \frac{\alpha}{2} \left\| \mathbf{Z}^{(v)} \right\|_F^2 + \frac{\rho}{2} \sum_{v=1}^m \left\| \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} + \frac{1}{\rho} \mathbf{Y}^{(v)} \right\|_F^2 + \\ & \frac{\beta}{2} \sum_{i=1}^n \left\| \mathbf{s}_i^T - \mathbf{w}_i^T \tilde{\mathbf{J}}_i \right\|_2^2 + \frac{\gamma}{2} \left\| \mathbf{S} \right\|_F^2 \quad (9) \\ & \text{s.t. } \mathbf{Z}^{(v)} \geq 0, \mathbf{S} \geq 0, \mathbf{w}_i \geq 0, \mathbf{S} \mathbf{1} = \mathbf{1}, \mathbf{w}_i^T \mathbf{1} = \mathbf{1}, \end{aligned}$$

where  $\rho$  is penalty term and  $\mathbf{Y}^{(v)}$  is Lagrangian multiplier.

#### Update $\mathbf{Z}^{(v)}$ :

Fixing all the variables except  $\mathbf{Z}^{(v)}$ , Eq. (6) can be written as follow:

$$\begin{aligned} & \min_{\mathbf{Z}^{(v)}} \sum_{v=1}^m \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} \right\|_F^2 + \frac{\alpha}{2} \left\| \mathbf{Z}^{(v)} \right\|_F^2 + \\ & \frac{\rho}{2} \sum_{v=1}^m \left\| \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} + \frac{1}{\rho} \mathbf{Y}^{(v)} \right\|_F^2 \quad (10) \\ & \text{s.t. } \mathbf{Z}^{(v)} \geq 0. \end{aligned}$$

Taking the derivative of Eq. (10) and setting it to zero, we have:

$$\begin{aligned} \mathbf{Z}^{(v)} &= \left( 2 \left( \mathbf{X}^{(v)} \right)^T \mathbf{X}^{(v)} + (\alpha + \rho) \mathbf{I} \right)^{-1} \\ & \left( 2 \left( \mathbf{X}^{(v)} \right)^T \mathbf{X}^{(v)} + \rho \mathbf{J}^{(v)} - \mathbf{Y}^{(v)} \right) \quad (11) \end{aligned}$$

#### Update $\mathbf{J}^{(v)}$ :

In order to solve  $\mathbf{J}^{(v)}$ , the Eq. (6) reduces to:

$$\begin{aligned} & \min_{\mathbf{J}^{(v)}} \frac{\rho}{2} \sum_{v=1}^m \left\| \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} + \frac{1}{\rho} \mathbf{Y}^{(v)} \right\|_F^2 + \\ & \frac{\beta}{2} \sum_{i=1}^n \left\| \mathbf{s}_i^T - \mathbf{w}_i^T \tilde{\mathbf{J}}_i \right\|_2^2, \quad (13) \end{aligned}$$

---

#### Algorithm 1: Augmented Lagrangian MethOd fiNe-graineD(ALMOND) Multi-view Clustering

---

**Require:** Multi-view data  $\mathbf{X}^{(v)}$ . ( $v = 1, 2, \dots, m$ ) Parameter  $\alpha, \beta$  and  $\gamma$ .

**Ensure:** Clustering result.

- 1: **repeat**
- 2:   Update  $\mathbf{Z}^{(v)}$  according to Eq. (11).
- 3:   Update  $\mathbf{J}^{(v)}$  by solving Eq. (15).
- 4:   Update  $\mathbf{W}$  according to Eq. (18).
- 5:   Update  $\mathbf{S}$  according to Eq. (21).
- 6:   Update Lagrangian multiplier  $\mathbf{Y}^{(v)}$  by:

$$\mathbf{Y}_{t+1}^{(v)} = \mathbf{Y}_t^{(v)} + \rho \left( \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} \right) \quad (12)$$

- 7: **until** converge
  - 8: Conduct the standard spectral clustering on the optimal graph  $\mathbf{S}$  to obtain the final clustering result.
- 

which obviously can be rewritten in a sample-wise form as:

$$\min_{\mathbf{J}^{(v)}} \frac{\rho}{2} \sum_{i=1}^n \left\| \tilde{\mathbf{Z}}_i - \tilde{\mathbf{J}}_i + \frac{1}{\rho} \tilde{\mathbf{Y}}_i \right\|_F^2 + \frac{\beta}{2} \left\| \mathbf{s}_i^T - \mathbf{w}_i^T \tilde{\mathbf{J}}_i \right\|_2^2. \quad (14)$$

We take the first order derivative of  $\tilde{\mathbf{J}}_i$ . Therefore we have:

$$\tilde{\mathbf{J}}_i = (\beta \mathbf{w}_i \mathbf{w}_i^T + \rho \mathbf{I})^{-1} \left( \beta \mathbf{w}_i \mathbf{s}_i^T + \rho \tilde{\mathbf{Z}}_i + \tilde{\mathbf{Y}}_i \right) \quad (15)$$

#### Update $\mathbf{W}$ :

As we optimize the weight matrix  $\mathbf{W}$ , Eq. (6) becomes:

$$\begin{aligned} & \min_{\mathbf{W}} \frac{\beta}{2} \sum_{i=1}^n \left\| \mathbf{s}_i^T - \mathbf{w}_i^T \tilde{\mathbf{J}}_i \right\|_2^2 \quad (16) \\ & \text{s.t. } \mathbf{w}_i \geq 0, \mathbf{w}_i^T \mathbf{1} = \mathbf{1}. \end{aligned}$$

By letting  $\mathbf{K}_i = \left( \mathbf{1} \mathbf{s}_i^T - \tilde{\mathbf{J}}_i \right) \in \mathbb{R}^{m \times n}$ , for each sample, Eq. (16) can be simplified as:

$$\begin{aligned} & \min_{\mathbf{W}} \left\| \mathbf{w}_i^T \mathbf{K}_i \right\|_2^2 \quad (17) \\ & \text{s.t. } \mathbf{w}_i \geq 0, \mathbf{w}_i^T \mathbf{1} = \mathbf{1}. \end{aligned}$$

We can solve Eq. (17) by taking its derivative with respect to  $\mathbf{w}_i$  and setting it to zero:

$$\mathbf{w}_i = \frac{(\mathbf{K}_i \mathbf{K}_i^T)^{-1} \mathbf{1}}{\mathbf{1}^T (\mathbf{K}_i \mathbf{K}_i^T)^{-1} \mathbf{1}}. \quad (18)$$

#### Update $\mathbf{S}$ :

$$\begin{aligned} & \min_{\mathbf{S}} \frac{\beta}{2} \sum_{i=1}^n \left\| \mathbf{s}_i^T - \mathbf{w}_i^T \tilde{\mathbf{J}}_i \right\|_2^2 + \frac{\gamma}{2} \left\| \mathbf{S} \right\|_F^2 \quad (19) \\ & \text{s.t. } \mathbf{S} \geq 0, \mathbf{S} \mathbf{1} = \mathbf{1}, \end{aligned}$$

Method	ACC	NMI	Purity	F-score
AWP	61.82	65.77	61.82	49.71
CDG	47.88	55.70	49.15	37.61
mPAC	25.03	26.38	27.21	12.98
Co-reg	56.36	60.30	57.64	41.43
Co-train	53.94	57.90	55.52	38.68
FPMVS	44.24	49.76	46.67	30.45
LMVSC	61.45	61.96	68.24	40.57
MCGC	68.36	70.36	68.55	52.27
MLAN	67.27	68.92	67.27	48.33
MVCTM	70.30	71.54	70.30	<b>56.71</b>
MVGL	62.42	66.14	62.42	47.40
SwMC	58.36	64.82	71.58	46.90
CoMSC	63.58	65.17	70.36	44.35
COMVSC	71.27	71.70	71.33	55.01
Ours	<b>71.52</b>	<b>72.37</b>	<b>71.64</b>	56.40

Table 2: The clustering results on Yale.

Method	ACC	NMI	Purity	F-score
AWP	26.96	6.86	28.50	29.55
CDG	33.36	10.79	33.85	31.93
mPAC	37.45	18.21	38.17	31.63
CoMSC	60.37	33.66	61.34	44.42
COMVSC	34.12	7.38	34.67	28.24
Co-reg	35.08	16.14	40.54	28.42
Co-train	52.97	26.91	55.41	38.62
FPMVS	38.06	14.72	42.82	27.89
LMVSC	49.88	24.39	52.96	35.17
MCGC	42.40	21.96	44.80	33.27
MLAN	31.46	5.60	33.00	24.76
MVCTM	51.42	26.67	53.93	37.04
MVGL	28.11	6.41	28.32	30.04
SwMC	40.18	14.34	52.26	29.20
Ours	<b>62.43</b>	<b>35.71</b>	<b>65.66</b>	<b>46.92</b>

Table 3: Clustering performance on dataset Citeseer.

which can be reformulated as:

$$\begin{aligned} \min_{\mathbf{s}_i} (\beta + \gamma) \mathbf{s}_i \mathbf{s}_i^T - 2\beta \mathbf{w}_i^T \tilde{\mathbf{J}}_i \mathbf{s}_i \\ \text{s.t. } \mathbf{s}_i \geq 0, \mathbf{s}_i \mathbf{1} = 1. \end{aligned} \quad (20)$$

Thus, we can derive a close form of Eq. (20):

$$\begin{aligned} \min_{\mathbf{s}_i} \left\| \mathbf{s}_i - \frac{\beta}{\beta + \gamma} \mathbf{w}_i^T \tilde{\mathbf{J}}_i \right\|_2^2 \\ \text{s.t. } \mathbf{s}_i \geq 0, \mathbf{s}_i \mathbf{1} = 1. \end{aligned} \quad (21)$$

It is obvious that Eq. (21) can be tackled by the optimization algorithm proposed in (Huang, Nie, and Huang 2015).

## Experiment

In the following passage, we present empirical evidence showcasing the efficacy and superiority of our proposed methodology on a variety of benchmark datasets. This is achieved through a comprehensive comparison with alternative state-of-the-art multi-view clustering techniques.

Method	ACC	NMI	Purity	F-score
AWP	30.76	14.27	38.96	28.17
CDG	43.94	21.02	46.42	35.86
mPAC	45.77	24.45	49.40	34.34
Co-reg	35.41	20.05	43.37	27.44
Co-train	50.95	31.32	56.45	36.64
FPMVS	64.84	40.23	64.84	45.09
LMVSC	43.40	25.33	49.14	31.59
MCGC	48.88	32.89	54.40	36.27
MLAN	32.05	3.91	32.20	29.64
MVCTM	39.00	23.82	46.34	31.18
MVGL	41.29	16.80	43.32	33.14
SwMC	49.97	26.92	66.87	38.47
CoMSC	64.16	46.49	68.67	49.46
COMVSC	34.70	13.79	38.19	29.49
Ours	<b>73.00</b>	<b>53.93</b>	<b>73.31</b>	<b>59.81</b>

Table 4: The clustering results on Cora.

Method	ACC	NMI	Purity	F-score
AWP	66.32	43.43	66.32	52.69
CDG	90.06	76.40	90.06	82.65
mPAC	89.74	78.34	89.74	83.51
Co-reg	71.44	51.12	71.44	57.26
Co-train	84.24	66.36	84.68	73.80
FPMVS	48.11	20.75	49.76	38.44
LMVSC	59.65	39.75	66.86	44.08
MCGC	87.54	68.85	87.54	77.56
MLAN	31.86	7.03	32.41	34.67
MVCTM	65.69	47.24	65.69	58.64
MVGL	70.58	58.87	70.58	66.38
SwMC	42.62	17.50	45.23	34.08
CoMSC	91.96	77.73	91.96	85.47
COMVSC	53.57	29.40	55.13	41.81
Ours	<b>94.83</b>	<b>84.08</b>	<b>94.83</b>	<b>89.98</b>

Table 5: Clustering performance on dataset BBC.

Method	ACC	NMI	Purity	F-score
AWP	74.50	86.02	76.25	65.18
CDG	57.80	73.12	63.42	31.59
mPAC	36.60	53.17	40.68	14.73
Co-reg	61.65	79.38	66.22	51.39
Co-train	61.90	79.23	65.60	50.88
FPMVS	55.45	73.72	59.27	41.02
LMVSC	65.65	80.37	75.00	49.64
MCGC	79.62	90.49	83.42	73.99
MLAN	77.00	87.22	82.75	56.25
MVCTM	63.25	82.47	66.50	57.90
MVGL	59.50	72.58	66.75	23.98
SwMC	61.90	78.73	68.83	49.39
CoMSC	70.15	83.95	76.85	58.98
COMVSC	78.83	89.10	83.42	71.52
Ours	<b>81.75</b>	<b>91.57</b>	<b>85.68</b>	<b>77.10</b>

Table 6: Clustering performance on dataset ORL.

## Experiment Setup

As for benchmark datasets, we adopt **Yale**, **ORL**, **bbc-seg13of3**, **Cora**, **Cornell**, and **citeseer**. Specifically, **Yale**<sup>1</sup>

<sup>1</sup><https://www.kaggle.com/datasets/olgabelitskaya/yale-face-database>

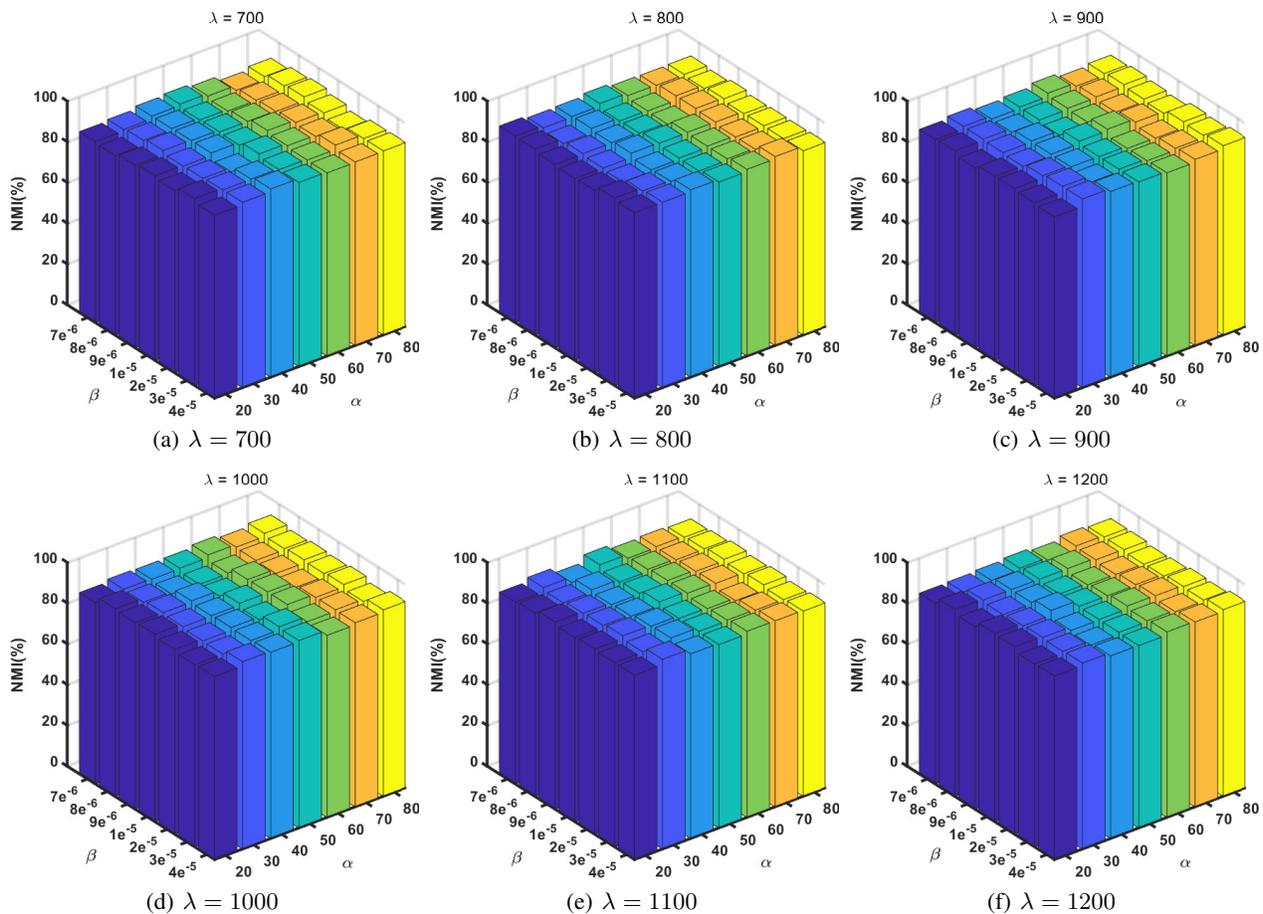


Figure 2: parameter analysis of our algorithm on ORL dataset.

database contains face images of 15 subjects with different conditions images in raw pixel. The Olivetti Research Laboratory **ORL**<sup>2</sup> face data set consists of 400 face images in 40 different themes in total. For each subject, the images are described in three features: facial expressions, facial details, and lighting. **bbcseg13of3**<sup>3</sup> is a subset of BBC data set, which consists of documents from the BBC news website corresponding to stories in five topical areas (business, entertainment, politics, sport, tech). **Cora**<sup>4</sup> data set consists of 2708 scientific publications classified into one of seven classes. **Cornell** is a subset of WebKB<sup>5</sup>, which is collected by Cornell University, contains 195 web pages and a web page is made of 2 views: content features and cites features. **citeseer**<sup>6</sup> is the citation network extracted from the CiteSeer digital library, which consists of 3312 scientific publications classified into one of six classes.

With the aim of conducting a thorough assessment, we

<sup>2</sup><https://www.kaggle.com/datasets/tavarez/the-orl-database-for-training-and-testing>

<sup>3</sup><http://mlg.ucd.ie/datasets/bbc.html>

<sup>4</sup><https://relational.fit.cvut.cz/dataset/CORA>

<sup>5</sup><https://starling.utdallas.edu/datasets/webkb/>

<sup>6</sup><https://relational.fit.cvut.cz/dataset/CiteSeer>

undertake a comparative analysis of our proposed method against several competing approaches: Multiview Clustering via Adaptively Weighted Procrustes (AWP) (Nie, Tian, and Li 2018), Multi-view Clustering via Cross-view Graph Diffusion (CGD) (Tang et al. 2020), Multiple Partitions Aligned Clustering (mPAC) (Kang et al. 2019), Multiview Subspace Clustering via Co-training (CoMSC) (Liu et al. 2021), Consensus One-step Multi-view Subspace Clustering (COMVSC) (Zhang et al. 2020), Multi-view Spectral clustering with Co-reg strategy (Co-reg) (Kumar and Daumé 2011), Multi-view Spectral Clustering with Co-train strategy (Co-train) (Kumar and Daumé 2011), Fast parameter-free multi-view subspace clustering (FPMVS) (Wang et al. 2021), (LMVSC) (), Multi-view Consensus Graph Clustering (MCGC) (Zhan et al. 2018), Multi-View Clustering with Adaptive Neighbours (MLAN) (Nie, Cai, and Li 2017), Multi-view Clustering on Topological Manifold (MVCTM) (Huang et al. 2022a), Self-weighted Multiview Clustering (SwMC) (Nie et al. 2017).

## Results Analysis

Four criteria were adopted in our experiments to validate the effectiveness and superiority of our proposed approach.

Method	ACC	NMI	Purity	F-score
AWP	38.97	13.63	49.74	33.72
CDG	51.33	16.72	56.15	45.37
mPAC	50.77	24.67	56.92	44.86
Co-train	41.38	22.36	54.26	34.86
Co-reg	36.26	13.68	47.28	31.98
MLAN	59.28	20.94	59.28	49.84
SwMC	46.67	13.82	52.82	41.51
MCGC	36.41	22.60	53.85	33.79
LRMSC	34.77	8.51	43.95	28.99
SMVSC	48.67	20.61	53.90	42.14
FPMVS	35.90	17.67	53.33	33.20
LMVSC	<b>56.72</b>	30.05	66.62	50.31
CoMSC	63.59	38.75	63.59	<b>57.05</b>
COMVSC	50.51	13.42	53.38	43.90
Ours	56.62	<b>44.64</b>	<b>73.74</b>	52.70

Table 7: Clustering performance on dataset Cornell.

They are Normalized Mutual Information (NMI), Accuracy (ACC), Purity, and F-Score. The experiment results are displayed in Tables.2-7. Note that, our method outperforms other competitors in most cases. Especially on Cora dataset, our approach has a significant advantage over the second-best method by 13.79%, 16.00%, 6.76%, and 20.93%, with respect to ACC, NMI, Purity, and F-Score. Meanwhile, our method makes an outstanding improvement on Cite-seer dataset as well, where the ACC, NMI, Purity, and F-Score are improved by 3.41%, 6.09%, 7.04%, and 5.63%. In BBC dataset, 3.12%, 8.17%, 3.12%, and 5.28% improvements have been made with respect to ACC, NMI, Purity, and F-Score. Whereas, in terms of F-Score on Yale dataset, MVCTM has the best result, On Cornell dataset, LMVSC and CoMSC slightly surpassed our method in terms of ACC and F-Score respectively. Generally, the empirical study has proven the advantage of our ALMOND for multi-view clustering over other state-of-the-art.

### Parameter Analysis

With the purpose of examining how different parameter settings will affect the results of clustering, we change the values of  $\alpha$ ,  $\beta$  and  $\gamma$  in the ranges  $[30, \dots, 80]$ ,  $[7e^{-6}, \dots, 4e^{-5}]$ , and  $[700, \dots, 1200]$ , respectively. Taking the ORL dataset as an example, we can see the clustering performance is quite stable with respect to different parameter settings within a certain range, as shown in Fig. (2). But in our extensive experiments, we have also found that the clustering performance is a little unstable when the parameters are chosen in a wider range. This instability is highly likely brought by the Lagrangian parameter  $\rho$ . For simplicity, we set it as 1 from the beginning and didn't consider it as a hyper-parameter to search in our experiment. But it grows exponentially in every iteration according to the ALM principle, i.e.,  $\rho \leftarrow 1.2\rho$ . This is the possible reason why the clustering performance is a little unstable with respect to a broader parameter settings. In our future work, we will try different schemes to inherently overcome this problem, such as graph fusion based on adaptive learning method.

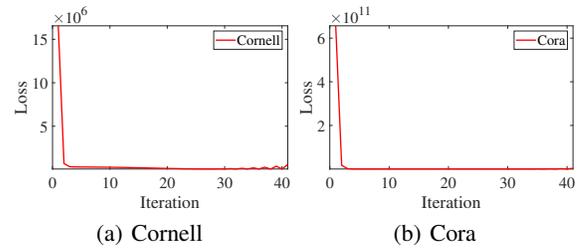


Figure 3: Convergence curve of our algorithm.

### Convergence Analysis

Due to the intrinsic non-convex nature of the optimization involved in our proposed method, which necessitates the utilization of an iterative algorithm for its resolution, it becomes imperative to meticulously ascertain the convergence properties exhibited by our model. Therefore, this section empirically showcases the convergence property and how fast our algorithm can converge. As shown in Fig. (3), our optimization framework can easily find a globally optimal solution within several iterations, which has indicated the convergence property of our method. However, we also need to point out that at the early age of iterations, the loss of our objective function is not very stable and does not drop smoothly. An intuitive reason for this might be the conflict between capturing the global information by subspace learning and exploiting the local structure by fine-grained weighting strategy.

### Conclusion

In this paper, we proposed an effective Augmented Lagrangian Method for fine-grainD (ALMOND) multi-view optimization. It is distinctive from traditional ways of weighing the importance of different views on the view level and self-consistent. Based on the assumption that the consistency among multi-view information mainly lay in the generally same local structure, our strategy focuses on the intersections of different views sample-wisely. Meanwhile, self-consistency is better maintained in our model, where self-consistency is achieved by developing an effective ALMOND multi-view optimization. Therefore, we overcame the problem of no interpretability that exists in previous work. Furthermore, our method generates outstanding results on several authoritative benchmark datasets and is proven to outperform current state-of-the-art methods. In the future, we will study the relationship and conflict between global structure retrieval and local information preservation, and how they will influence the convergence property of the objective function.

### Acknowledgements

This work was partially supported by the National Science Foundation of China under Grant 62106164 and 62376175, the State Key Program of the National Science Foundation of China under Grant 61836006, the 111 Project under Grant B21044, and the Sichuan Science and Technology Program under Grants 2021ZDZX0011.

## References

- Andreani, R.; Birgin, E. G.; Martínez, J. M.; and Schuverdt, M. L. 2008. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4): 1286–1309.
- Dalal, N.; and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, volume 1, 886–893.
- 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.
- Huang, J.; Nie, F.; and Huang, H. 2015. A new simplex sparse learning model to measure data similarity for clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3569–3575.
- Huang, S.; Tsang, I.; Xu, Z.; Lv, J.; and Liu, Q.-H. 2022a. Multi-View Clustering on Topological Manifold. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 6944–6951.
- Huang, S.; Tsang, I. W.; Xu, Z.; and Lv, J. 2021. Measuring diversity in graph learning: a unified framework for structured multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(12): 5869–5883.
- Huang, S.; Tsang, I. W.; Xu, Z.; and Lv, J. 2022b. CGDD: Multiview Graph Clustering via Cross-Graph Diversity Detection. *IEEE Transactions on Neural Networks and Learning Systems*.
- Huang, X.; Cheng, X.; Geng, Q.; Cao, B.; Zhou, D.; Wang, P.; Lin, Y.; and Yang, R. 2018. The ApolloScape Dataset for Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Kang, Z.; Guo, Z.; Huang, S.; Wang, S.; Chen, W.; Su, Y.; and Xu, Z. 2019. Multiple partitions aligned clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2701–2707.
- Kumar, A.; and Daumé, H. 2011. A co-training approach for multi-view spectral clustering. In *Proceedings of the International Conference on Machine Learning*, 393–400.
- Ling, Y.; Chen, J.; Ren, Y.; Pu, X.; Xu, J.; Zhu, X.; and He, L. 2023. Dual Label-Guided Graph Refinement for Multi-View Graph Clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(7): 8791–8798.
- Liu, J.; Liu, X.; Yang, Y.; Guo, X.; Kloft, M.; and He, L. 2021. Multiview subspace clustering via co-training robust data representation. *IEEE Transactions on Neural Networks and Learning Systems*.
- Liu, J.; Liu, X.; Yang, Y.; Liao, Q.; and Xia, Y. 2023. Contrastive Multi-View Kernel Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Liu, S.; Wang, S.; Zhang, P.; Xu, K.; Liu, X.; Zhang, C.; and Gao, F. 2022. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7576–7584.
- Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60: 91–110.
- Nie, F.; Cai, G.; and Li, X. 2017. Multi-view clustering and semi-supervised classification with adaptive neighbours. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Nie, F.; Li, J.; and Li, X. 2016. Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1881–1887.
- Nie, F.; Li, J.; Li, X.; et al. 2017. Self-weighted Multi-view Clustering with Multiple Graphs. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2564–2570.
- Nie, F.; Tian, L.; and Li, X. 2018. Multiview clustering via adaptively weighted procrustes. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2022–2030.
- Ojala, T.; Pietikainen, M.; and Maenpaa, T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7): 971–987.
- Pan, E.; and Zhao, K. 2021. Multi-view Contrastive Graph Clustering. In *Advances in Neural Information Processing Systems*, volume 34, 2148–2159.
- Siwei, W.; Liu, X.; Suyuan, L.; Jiaqi, J.; Tu, W.; Zhu, X.; and En, Z. 2022. Align then Fusion: Generalized Large-scale Multi-view Clustering with Anchor Matching Correspondences. In *Advances in Neural Information Processing Systems*, volume 35, 5882–5895.
- Tan, Y.; Liu, Y.; Huang, S.; Feng, W.; and Lv, J. 2023a. Sample-Level Multi-View Graph Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23966–23975.
- Tan, Y.; Liu, Y.; Wu, H.; Lv, J.; and Huang, S. 2023b. Metric Multi-View Graph Clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(8): 9962–9970.
- Tang, C.; Liu, X.; Zhu, X.; Zhu, E.; Luo, Z.; Wang, L.; and Gao, W. 2020. CGD: Multi-view clustering via cross-view graph diffusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 5924–5931.
- Wang, H.; Yang, Y.; and Liu, B. 2019. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6): 1116–1129.
- Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2021. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31: 556–568.
- Wang, Y.; Pei, X.; and Zhan, H. 2022. Fine-grained Graph Learning for Multi-view Subspace Clustering. *arXiv preprint arXiv:2201.04604*.
- Wang, Z.; Chen, L.; Zhang, J.; Yin, Y.; and Li, D. 2020. Multi-view ensemble learning with empirical kernel for heart failure mortality prediction. *International Journal*

for *Numerical Methods in Biomedical Engineering*, 36(1): e3273.

Xia, W.; Gao, Q.; Wang, Q.; Gao, X.; Ding, C.; and Tao, D. 2023. Tensorized Bipartite Graph Learning for Multi-View Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 5187–5202.

Yu, X.; Liu, H.; Lin, Y.; Liu, N.; and Sun, S. 2023. Sample-level weights learning for multi-view clustering on spectral rotation. *Information Sciences*, 619: 38–51.

Zhan, K.; Nie, F.; Wang, J.; and Yang, Y. 2018. Multiview consensus graph clustering. *IEEE Transactions on Image Processing*, 28(3): 1261–1270.

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*.