# Towards Resource-friendly, Extensible and Stable Incomplete Multi-view Clustering

**Shengju Yu** [1] **Zhibing Dong** [1] **Siwei Wang** [2] **Xinhang Wan** [1]
**Yue Liu** [1] **Weixuan Liang** [1] **Pei Zhang** [1] **Wenxuan Tu** [1] **Xinwang Liu** [1]

## Abstract

Incomplete multi-view clustering (IMVC) methods typically encounter three drawbacks: (1) intense time and/or space overheads; (2) intractable hyper-parameters; (3) non-zero variance results. With these concerns in mind, we give a simple yet effective IMVC scheme, termed as ToRES. Concretely, instead of self-expression affinity, we manage to construct prototype-sample affinity for incomplete data so as to decrease the memory requirements. To eliminate hyper-parameters, besides mining complementary features among views by view-wise prototypes, we also attempt to devise cross-view prototypes to capture consensus features for jointly forming worth-having clustering representation. To avoid the variance, we successfully unify representation learning and clustering operation, and directly optimize the discrete cluster indicators from incomplete data. Then, for the resulting objective function, we provide two equivalent solutions from perspectives of feasible region partitioning and objective transformation. Many results suggest that ToRES exhibits advantages against 20 SOTA algorithms, even in scenarios with a higher ratio of incomplete data.

## 1. Introduction

Multi-view data (MVD), generally collected from diverse domains or modalities, is becoming increasingly common in daily life. As a technique for handling MVD, multi-view clustering (MVC) can partition it into distinct groups without requiring any prior label information (Xu et al., 2022; Yang et al., 2021; Yu et al., 2023a; 2024a; Xia et al., 2022b; Fu et al., 2024). However, due to transmission errors or privacy protection, incomplete instances are unavoidable, which leads to existing MVC methods not working properly. Accordingly, incomplete multi-view clustering (IMVC) grasps widespread attention recently (Peng et al., 2019; Liu et al., 2021; Wan et al., 2024b; Liang et al., 2023).

Given incomplete data $\{\mathbf{D}_m \in \mathbb{R}^{d_m \times n}\}_{m=1}^M$ and index vectors $\{w_m \in \mathbf{R}^{n_m}\}_{m=1}^M$ where $n_m$ and $d_m$ represent the number and dimension of samples observed on view $m$, one can construct indicator matrices $\{\mathbf{W}_m \in \mathbb{R}^{n \times n_m}\}_{m=1}^M$ in which $[\mathbf{W}_m]_{i,j} = 1 \ \ if \ \ [w_m]_j = i \ \ else \ \ [\mathbf{W}_m]_{i,j} = 0, \ \forall j = 1, 2, \cdots, n_m$. Then, $\mathbf{D}_m \mathbf{W}_m \in \mathbb{R}^{d_m \times n_m}$ represents available instances on view $m$. On this basis, current methods first build self-expression affinity on each view by

$$\begin{aligned} \min_{\mathbf{H}_m} \sum_{m=1}^{M} \|\mathbf{D}_m \mathbf{W}_m - \mathbf{D}_m \mathbf{W}_m \mathbf{H}_m\|_F^2 + \lambda \phi(\mathbf{H}_m) \\ \text{s.t. } \mathbf{H}_m \in \mathbb{R}^{n_m \times n_m} \geq 0, \mathbf{H}_m^\top \mathbf{1} = \mathbf{1}. \end{aligned} \tag{1}$$

Afterwards, the full affinity is restored via $\mathbf{W}_m \mathbf{H}_m \mathbf{W}_m^\top \in \mathbb{R}^{n \times n}$. Due to missing instances, there will be blank rows and columns in the full affinity. By locating these affinities with blanks into corresponding positions, current methods optimize a set of sketched affinities to generate the fused full affinity $\mathbf{H}$:

$$\begin{aligned} \min_{\mathbf{H}} \psi \left(\mathbf{H}, \mathbf{W}_1 \mathbf{H}_1 \mathbf{W}_1^\top, \cdots, \mathbf{W}_M \mathbf{H}_M \mathbf{W}_M^\top; \beta\right) \\ \text{s.t. } \mathbf{H} \in \mathbb{R}^{n \times n} \geq 0, \mathbf{H}^\top \mathbf{1} = \mathbf{1}. \end{aligned} \tag{2}$$

Subsequently, running spectral clustering on $\mathbf{H}$ generates discrete data labels. Here, $\phi$ and $\psi$ denote regularization and fusion strategies respectively. $\lambda$ and $\beta$ are hyper-parameters.

Although making promising results, the paradigm adopted by current methods typically requires at least $\mathcal{O}(n^3)$ computing cost and $\mathcal{O}(n^2)$ memory cost due to both $\mathbf{W}_m \mathbf{H}_m \mathbf{W}_m^\top$ and $\mathbf{H}$ being $\mathbb{R}^{n \times n}$. Moreover, the presence of hyper-parameters harms the extensibility of models. Even worse, the generated results usually contain variance, which not only decreases the reliability of clustering but also leads to relatively poor interpretability. To alleviate these three

[1]School of Computer, National University of Defense Technology, Changsha, Hunan, China [2]Intelligent Game and Decision Lab, Beijing, China. yu-shengju@foxmail.com. Correspondence to: Siwei Wang <wangsiwei13@nudt.edu.cn>, Xinwang Liu <xinwangliu@nudt.edu.cn>.

*Proceedings of the 41st International Conference on Machine Learning*, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

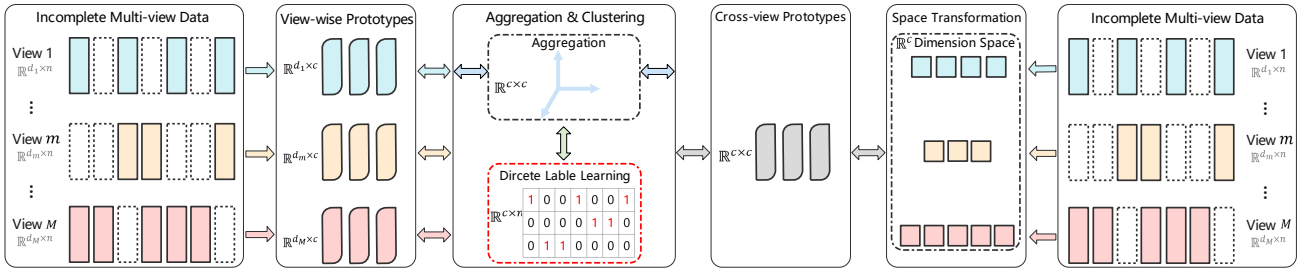



*Figure 1.* Framework of ToRES. It builds prototype-sample affinity with small size for incomplete views, and also does not involve the fusion stage like in Eq. (2). All prototype information is gathered via one aggregation matrix. To form desirable representation under without the help of hyper-parameters, it designs two types of prototypes, view-wise and cross-view, to jointly explore multi-view data features. To output stable results, it concurrently learns representation and performs clustering, and optimizes the cluster labels directly.

shortcomings, we devise a ToRES algorithm, and its framework is shown in Fig. 1. Concretely, rather than building self-expression affinity, we seek to learn the prototypes for each incomplete view, and build prototype-sample affinity with small size, thereby decreasing the complexity. We also skip the fusion stage by directly gathering all prototype information using one aggregation matrix. Besides, combined with Eqs. (1) and (2), we hold that hyper-parameters aim at balancing the error term and regularization term so as to form desirable clustering representation. Inspired by this, we introduce two types of prototypes for incomplete data to jointly explore multi-view features. Cross-view prototypes are devoted to capturing consensus information among views while view-wise prototypes capture complementary information. To avoid variance, we integrate representation learning and clustering operation together, and directly optimize the discrete labels from incomplete data, which not only well preserves the original diversity of samples but also generates stable results, decreasing the fuzziness. Then, for minimizing the resulting objective, we give two equivalent solutions from perspectives of feasible region partitioning and objective transformation. Further, we organize a series of experiments on datasets with multiple scales and missing ratios to demonstrate the effectiveness of ToRES sufficiently. In short, contributions in this work are as follows:

**1)** We design an innovative IMVC method, called ToRES. To reduce complexity, it successfully builds prototype-sample affinity for incomplete views instead of self-expression one. The characteristic of friendly to computing and memory resources makes ToRES able to handle large-scale problems.

**2)** To get rid of hyper-parameters, ToRES introduces view-wise and cross-view prototypes for incomplete data to jointly exploit multi-view features. The characteristic of no hyper-parameters makes ToRES easily extensible to other scenes.

**3)** To eliminate variance, ToRES unifies representation learning and clustering operation, and directly learns the discrete labels from incomplete data. The characteristic of zero variance makes ToRES own the ability to output stable results.

**4)** To minimize the resulting objective function, ToRES provides two optimization schemes. Experiments conducted on several datasets with different scales and incomplete percentages validate ToRES's effectiveness and strong points.

## 2. Related Work

We roughly classify existing IMVC algorithms into four categories: kernel-style (Li et al., 2021b; Zhang et al., 2021; Liu et al., 2020), NMF-style (Zhao et al., 2016; Xia et al., 2022a; Wen et al., 2021a), graph-style (Yu et al., 2024b; Li et al., 2021a; Chen et al., 2021; Wen et al., 2021b; Yu et al., 2023b) and deep learning style (Lin et al., 2021; Liu et al., 2023; Wang et al., 2021). Kernel-style methods seek to generate the kernel with consistent structures among all partial views. For instance, Li et al. (2021b) build up the local and global kernel alignment standard, and introduce matrix-induced regularizer to improve the correlation between kernels. NMF-style methods aim to construct consensus representation from partial views using adaptive view-specific weights. For instance, Wen et al. (2021a) learn the local information and semantic consistency concurrently, and impose adaptive weights to measure the contributions of different views. Graph-style methods attempt to reconstruct the similarity graph with compatible characteristics from multiple partial views. For example, Li et al. (2021a) jointly learn the partition space and consensus graph, and utilize a unified partition indicator to uncover the similarity between incomplete data. Deep learning style methods are devoted to exploiting higher-order information from partial views. For example, Xue et al. (2022) employ a contrastive regularizer to exploit data correlation, and adopt a specific encoder network to alleviate the impact of partial views. Despite producing pleasing results, they generally suffer from intensive overheads, hyper-parameters and non-zero variances, which severely limits their further deployment.

## 3. Methodology

### 3.1. Designed Framework

Denote $\mathbf{G}_m \in \mathbb{R}^{c\times d_m}$, $\mathbf{D}_m \in \mathbb{R}^{d_m\times n}$, $\mathbf{W}_m \in \mathbb{R}^{n\times n_m}$, $\mathbf{E} \in \mathbb{R}^{c\times c}$, $\mathbf{O} \in \mathbb{R}^{c\times c}$, $\mathbf{L} \in \mathbb{R}^{c\times n}$ and $\mathbf{E}_m \in \mathbb{R}^{d_m\times c}$ as space transformation matrix, original data matrix, indicator matrix, cross-view prototype matrix, aggregation matrix, discrete cluster label matrix and view-wise prototype matrix, respectively. $d_m$, $c$ and $n$ represent the data dimension on view $m$, the number of clusters and samples respectively. Then, according to Fig. 1, our ToRES is formulated as

$$\begin{aligned}\min_{\mathbf{G}_m,\mathbf{E},\mathbf{O},\mathbf{L},\mathbf{E}_m} &\sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 \\ &+ \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2 \\ \text{s.t. } &\mathbf{G}_m\mathbf{G}_m^\top = \mathbf{I}_c, \mathbf{O}^\top\mathbf{O} = \mathbf{I}_c, \\ &\mathbf{L} \in \{0,1\}^{c\times n}, \|\mathbf{L}_{:,j}\|_1 = 1, j \in \{1,2,\dots,n\},\end{aligned} \tag{3}$$

where $\mathbf{G}_m$ aims to transform original data to a potential common space, and $\mathbf{O}$ gathers all prototype information.

### 3.2. Optimization

We adopt the idea of alternating optimization to solve Eq.(3).

★ ***Step-One:*** *Optimizing* $\mathbf{O}$. Fixing variables $\mathbf{G}_m$, $\mathbf{E}$, $\mathbf{L}$ and $\mathbf{E}_m$, Eq. (3) becomes

$$\begin{aligned}\min_{\mathbf{O}^\top\mathbf{O}=\mathbf{I}_c} &\sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 \\ &+ \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2 .\end{aligned} \tag{4}$$

We provide two schemes for solving Eq. (4).

**Scheme 1:**

For the first term in Eq. (4), we have

$$\begin{aligned}&\min_{\mathbf{O}} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 \Leftrightarrow \\ &\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}^\top\mathbf{EO}[\mathbf{A}_m]_{:,j} - 2[\mathbf{B}_m]_{:,j}^\top\mathbf{O}_{:,j},\end{aligned} \tag{5}$$

where $\mathbf{A}_m = \mathbf{LW}_m\mathbf{W}_m^\top\mathbf{L}^\top$, $\mathbf{B}_m = \mathbf{E}^\top\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top$.
To solve Eq. (5), we first give the following theorem:

**Theorem 3.1.** $\mathbf{A}_m$ *is a diagonal matrix.*

Based on **Theorem 3.1**, we have

$$\begin{aligned}&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}^\top\mathbf{EO}[\mathbf{A}_m]_{:,j} - 2[\mathbf{B}_m]_{:,j}^\top\mathbf{O}_{:,j} \Leftrightarrow \\ &\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}^\top\mathbf{E}[\mathbf{A}_m]_{j,j}\mathbf{O}_{:,j} - 2[\mathbf{B}_m]_{:,j}^\top\mathbf{O}_{:,j}.\end{aligned} \tag{6}$$

**Algorithm 1** Scheme 1 for solving Eq. (4)

**Input**: $\mathbf{G}_m$, $\mathbf{E}$, $\mathbf{L}$, $\mathbf{E}_m$, $\mathbf{D}_m$, $\mathbf{W}_m$.
**Output**: $\mathbf{O}$.
Construct $\mathbf{A}_m$, $\mathbf{B}_m$, $\mathbf{C}_m$.
1: **for** $j = 1$ to $c$ **do**
2: Update $\mathbf{O}_{:,j}$ by Eq. (8).
3: **end for**

For the second term in Eq. (4), we have

$$\begin{aligned}&\min_{\mathbf{O}} \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2 \Leftrightarrow \\ &\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}_m^\top\mathbf{E}_m[\mathbf{A}_m]_{j,j}\mathbf{O}_{:,j} - 2[\mathbf{C}_m]_{:,j}^\top\mathbf{O}_{:,j},\end{aligned} \tag{7}$$

where $\mathbf{C}_m = \mathbf{E}_m^\top\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top$.

Combined with Eqs. (5), (6) and (7), we have that Eq. (4) can be transformed as

$$\begin{aligned}\min_{\mathbf{O}_{:,j}} &\mathbf{O}_{:,j}^\top\left(\mathbf{E}^\top\mathbf{E}\sum_{m=1}^{M}[\mathbf{A}_m]_{j,j} + \sum_{m=1}^{M}\mathbf{E}_m^\top\mathbf{E}_m[\mathbf{A}_m]_{j,j}\right)\mathbf{O}_{:,j} \\ &+2\left(-\sum_{m=1}^{M}\mathbf{B}_m - \sum_{m=1}^{M}\mathbf{C}_m\right)_{:,j}^\top\mathbf{O}_{:,j} \\ \text{s.t. } &\frac{1}{2}\mathbf{O}_{:,j}^\top\mathbf{IO}_{:,j} + \left(-\frac{1}{2}\right) = 0, \\ &[\mathbf{O}_{:,1},\mathbf{O}_{:,2},\cdots,\mathbf{O}_{:,j-1},\mathbf{O}_{:,j+1}\cdots\mathbf{O}_{:,c}]^\top\mathbf{O}_{:,j} = \mathbf{0},\end{aligned} \tag{8}$$

which is a quadratic programming problem with quadratic constraints, and can be solved within $\mathcal{O}(c^3)$ complexity.

The detailed process of Scheme 1 is shown in **Algorithm 1**.

**Scheme 2:**

Unlike **Scheme 1** solving the problem from the perspective of feasible region partitioning, **Scheme 2** aims to transform the objective function.

For the objective function in Eq. (4), we have

$$\begin{aligned}&\text{Eq. (4)} \Leftrightarrow \\ &\max_{\mathbf{O}} \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{O}^\top\widetilde{\mathbf{F}}\mathbf{OA}_m + \mathbf{O}^\top\widetilde{\mathbf{F}}_m\mathbf{OA}_m + \mathbf{O}^\top\mathbf{S}_m\right),\end{aligned} \tag{9}$$

where $\widetilde{\mathbf{F}} = \lambda_{max}(\mathbf{F})\mathbf{I}_c - \mathbf{F}$, $\widetilde{\mathbf{F}}_m = \lambda_{max}(\mathbf{F}_m)\mathbf{I}_c - \mathbf{F}_m$. $\lambda_{max}(\cdot)$ represents the max eigenvalue. $\mathbf{F} = \mathbf{E}^\top\mathbf{E}$, $\mathbf{F}_m = \mathbf{E}_m^\top\mathbf{E}_m$, $\mathbf{S}_m = 2(\mathbf{B}_m + \mathbf{C}_m)$.

Denote $f(\mathbf{O}) = \sum_{m=1}^{M}\mathrm{Tr}(\mathbf{O}^\top\widetilde{\mathbf{F}}\mathbf{OA}_m + \mathbf{O}^\top\widetilde{\mathbf{F}}_m\mathbf{OA}_m + \mathbf{O}^\top\mathbf{S}_m)$, $\mathbf{O}_t$ and $\mathbf{O}_{t+1}$ as the variable values at the $t$-th and $(t+1)$-th iterations respectively. Then, we have

**Theorem 3.2.** *With* $\mathbf{O}_{t+1}$ *taking* $\mathbf{U}_t\mathbf{V}_t^\top$, *the following inequality holds*

$$\mathrm{Tr}\left(\mathbf{O}_{t+1}^\top\nabla f(\mathbf{O}_t)\right) \geq \mathrm{Tr}\left(\mathbf{O}_t^\top\nabla f(\mathbf{O}_t)\right), \tag{10}$$

**Algorithm 2** Scheme 2 for solving Eq. (4)

**Input**: $\mathbf{G}_m$, $\mathbf{E}$, $\mathbf{L}$, $\mathbf{E}_m$, $\mathbf{D}_m$, $\mathbf{W}_m$.
**Output**: $\mathbf{O}$.
Construct the function $f(\mathbf{O})$.
t=1.

1: **repeat**
2: Calculate $\nabla f(\mathbf{O}_t)$.
3: Perform SVD on $\nabla f(\mathbf{O}_t)$ to generate $\mathbf{U}_t$ and $\mathbf{V}_t^\top$.
4: $\mathbf{O}_{t+1} = \mathbf{U}_t\mathbf{V}_t^\top$.
5: t=t+1.
6: **until** $\|\mathbf{O}_t - \mathbf{O}_{t-1}\|_F / \|\mathbf{O}_{t-1}\|_F \leq 1e-5$.

*where $\nabla f(\cdot)$ denotes the gradient operation. $\mathbf{U}_t$ and $\mathbf{V}_t$ are the SVD results of $\nabla f(\mathbf{O}_t)$.*

Based on **Theorem 3.2**, we further have

**Theorem 3.3.** *$f(\mathbf{O})$ is non-decreasing under $\mathbf{O}_{t+1}$ taking $\mathbf{U}_t\mathbf{V}_t^\top$.*

Kindly note that in **Theorem 3.3**, $\mathbf{O}_{t+1}$ lies in the feasible region $\mathbf{O}^\top\mathbf{O} = \mathbf{I}_c$. Thus, we can determine the solution of $\mathbf{O}$ by comparing $\mathbf{O}_{t+1}$ and $\mathbf{O}_t$.

The detailed process of Scheme 2 is shown in **Algorithm** 2.

★ ***Step-Two:*** *Optimizing* $\mathbf{L}$. Fixing variables $\mathbf{G}_m$, $\mathbf{E}$, $\mathbf{O}$ and $\mathbf{E}_m$, Eq. (3) becomes

$$\begin{aligned} \min_{\mathbf{L}} \sum_{m=1}^{M} &\|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 \\ &+ \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2 \\ \text{s.t. } \mathbf{L} \in \{0,1\}^{c\times n}, &\|\mathbf{L}_{:,j}\|_1 = 1, j \in \{1,2,\ldots,n\}. \end{aligned} \tag{11}$$

Further, we have that Eq. (11) can be transformed as

$$\begin{aligned} &\min_{\mathbf{L}_{:,j}} \mathbf{L}_{:,j}^\top\mathbf{T}_j\mathbf{L}_{:,j} - \mathbf{J}_{j,:}\mathbf{L}_{:,j} \\ &\text{s.t. } \mathbf{L}_{:,j} \in \{0,1\}^{c\times 1}, \|\mathbf{L}_{:,j}\|_1 = 1, \end{aligned} \tag{12}$$

where

$$\begin{aligned} \mathbf{T}_j &= \mathbf{O}^\top \sum_{m=1}^{M} \left(\mathbf{E}^\top\mathbf{E}\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k} + \mathbf{E}_m^\top\mathbf{E}_m\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k}\right)\mathbf{O} \\ \mathbf{J} &= 2\sum_{m=1}^{M} \mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\left(\mathbf{G}_m^\top\mathbf{E} + \mathbf{E}_m\right)\mathbf{O}. \end{aligned} \tag{13}$$

The constraints mean that there is only one non-zero element in every column. Thus, we can determine $\mathbf{L}_{:,j}$ by comparing the diagonal elements of $\mathbf{T}_j$ and the $j$-th row of $\mathbf{J}$. That is,

$$i^* = \arg\min_i \ [\mathbf{T}_j]_{i,i} - \mathbf{J}_{j,i}, i \in \{1,2,\cdots,c\}. \tag{14}$$

When $i = i^*$, $\mathbf{L}_{i,j}$ is set as 1, otherwise 0.

**Algorithm 3** ToRES

**Input**: Original data $\{\mathbf{D}_m\}_{m=1}^M$, index vectors $\{w_m\}_{m=1}^M$.
**Output**: Discrete cluster label matrix $\mathbf{L}$.
**Initialize**: $\mathbf{O}$, $\mathbf{L}$, $\mathbf{G}_m$, $\mathbf{E}$, $\mathbf{E}_m$.
Construct indicator matrices $\{\mathbf{W}_m\}_{m=1}^M$.

1: **repeat**
2: Update $\mathbf{O}$ by Algorithm 1 or Algorithm 2.
3: Update $\mathbf{L}$ by Eq. (14).
4: Update $\mathbf{G}_m$ by Eq. (16).
5: Update $\mathbf{E}$ by Eq. (19).
6: Update $\mathbf{E}_m$ by Eq. (21).
7: **until** convergent

★ ***Step-Three:*** *Optimizing* $\mathbf{G}_m$. Fixing variables $\mathbf{E}$, $\mathbf{O}$, $\mathbf{L}$ and $\mathbf{E}_m$, Eq. (3) becomes

$$\min_{\mathbf{G}_m\mathbf{G}_m^\top = \mathbf{I}_c} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2. \tag{15}$$

Denote $\mathbf{P}_m = \mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top$, $\mathbf{Z}_m = 2\mathbf{EOLW}_m\mathbf{W}_m^\top\mathbf{D}_m^\top$, $\widetilde{\mathbf{P}}_m = \lambda_{max}(\mathbf{P}_m)\mathbf{I}_{d_m} - \mathbf{P}_m$. We have that the following theorem holds:

**Theorem 3.4.** *The solution of $\mathbf{G}_m$ can be obtained by*

$$\mathbf{G}_m = \mathbf{UV}^\top, \tag{16}$$

*where $\mathbf{U}$ and $\mathbf{V}$ are the SVD results of $\left(2[\mathbf{G}_m]_{pre}\widetilde{\mathbf{P}}_m + \mathbf{Z}_m\right)$. $[\mathbf{G}_m]_{pre}$ represents the value of $\mathbf{G}_m$ at previous step.*

★ ***Step-Four:*** *Optimizing* $\mathbf{E}$. Fixing variables $\mathbf{G}_m$, $\mathbf{O}$, $\mathbf{L}$ and $\mathbf{E}_m$, Eq. (3) becomes

$$\min_{\mathbf{E}} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2. \tag{17}$$

By setting its derivative to zero, we can get

$$\sum_{m=1}^{M} \left(\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\right)\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top = \mathbf{0}. \tag{18}$$

Based on the principles of multi-view clustering that sample appears on at least one view and that each cluster has at least one instance, we have that $\mathbf{A}_m = \mathbf{LW}_m\mathbf{W}_m^\top\mathbf{L}^\top$ is reversible. Thus, we have

$$\mathbf{E} = \sum_{m=1}^{M} \mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top \left(\mathbf{O}\sum_{m=1}^{M}\mathbf{A}_m\mathbf{O}^\top\right)^{-1} \tag{19}$$

★ ***Step-Five:*** *Optimizing* $\mathbf{E}_m$. Fixing variables $\mathbf{G}_m$, $\mathbf{E}$, $\mathbf{O}$ and $\mathbf{L}$, Eq. (3) becomes

$$\min_{\mathbf{E}_m} \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2. \tag{20}$$

*Table 1.* Average Clustering Results on Datasets Webkb, Wikifea and AWA10.

| Dataset | Methods | NoHp | 20% | | | 40% | | | 60% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | ACC | NMI | Purity | ACC | NMI | Purity | ACC | NMI | Purity |
| Webkb | IMSC-AGL | 3 | 50.18±0.05 | 0.46±0.01 | 78.12±0.00 | 60.89±0.00 | 0.29±0.00 | 78.12±0.00 | 50.09±0.05 | 0.20±0.00 | 78.12±0.00 |
| | AWP | 0 | 73.17±0.00 | 0.12±0.00 | 78.12±0.00 | 71.08±0.00 | 0.46±0.00 | 78.12±0.00 | 76.12±0.00 | 0.59±0.00 | 78.12±0.00 |
| | APMC | 2 | 77.55±0.00 | 0.50±0.00 | 78.12±0.00 | 80.40±0.00 | 11.25±0.00 | 80.40±0.00 | 70.01±0.00 | 23.49±0.00 | 70.01±0.00 |
| | IMG | 3 | 50.81±0.00 | 0.12±0.00 | 78.12±0.00 | 55.71±0.05 | 0.67±0.03 | 78.12±0.00 | 57.75±0.00 | 1.37±0.00 | 78.12±0.00 |
| | TMBSD | 2 | 54.04±0.00 | 0.82±0.00 | 78.12±0.00 | 52.14±0.00 | 0.99±0.00 | 78.12±0.00 | 55.59±0.06 | 0.12±0.00 | 78.12±0.00 |
| | IKMKC | 2 | 50.14±0.00 | 3.42±0.00 | 78.12±0.00 | 50.90±0.00 | 2.50±0.00 | 78.12±0.00 | 50.81±0.00 | 1.03±0.00 | 78.12±0.00 |
| | IMVTSC-MVI | 3 | 86.02±0.00 | 34.26±0.00 | 85.03±0.00 | 73.55±0.00 | 0.50±0.00 | 78.12±0.00 | 75.64±0.00 | 1.16±0.00 | 78.12±0.00 |
| | CPM-Nets | 1 | 56.14±0.13 | 0.01±0.00 | 78.12±0.02 | 69.80±0.88 | 0.49±0.13 | 78.12±0.23 | 73.32±0.61 | 1.34±0.14 | 78.40±0.17 |
| | LSIMVC | 4 | 67.36±0.00 | 1.01±0.00 | 78.12±0.00 | 69.27±0.00 | 0.20±0.00 | 78.12±0.00 | 71.84±0.02 | 0.26±0.00 | 78.12±0.00 |
| | GSRIMC | 3 | 65.56±0.00 | 0.18±0.00 | 78.12±0.00 | 55.09±0.00 | 0.16±0.00 | 78.12±0.00 | 55.57±0.00 | 0.13±0.00 | 78.12±0.00 |
| | COMPLETER | 3 | 62.28±0.70 | 6.04±0.70 | 78.12±0.92 | 61.84±0.54 | 5.38±0.05 | 78.12±0.79 | 64.74±0.60 | 5.85±0.30 | 78.12±0.25 |
| | TCIMC | 3 | 78.40±0.00 | 1.60±0.00 | 78.40±0.00 | 78.97±0.00 | 4.57±0.00 | 78.97±0.00 | 78.21±0.00 | 0.54±0.00 | 78.21±0.00 |
| | LRGR-IMVC | 2 | 68.79±0.00 | 3.43±0.00 | 78.12±0.00 | 72.22±0.00 | 1.55±0.00 | 78.12±0.00 | 61.37±0.00 | 2.27±0.00 | 78.12±0.00 |
| | BGIMVSC | 2 | 77.83±0.00 | 0.13±0.00 | 78.12±0.00 | 77.83±0.00 | 0.23±0.00 | 78.12±0.00 | 77.74±0.00 | 0.14±0.00 | 78.12±0.00 |
| | NGSP-CGL | 3 | 78.21±0.00 | 1.35±0.00 | 78.21±0.00 | 79.83±0.00 | 6.82±0.00 | 79.83±0.00 | 79.92±0.00 | 7.22±0.00 | 79.92±0.00 |
| | PIMVC | 2 | 82.11±0.00 | 15.56±0.09 | 82.11±0.00 | 76.02±0.00 | 11.11±0.00 | 78.12±0.00 | 77.93±0.00 | 17.54±0.00 | 78.12±0.00 |
| | ProImp | 2 | 67.46±0.45 | 16.34±0.56 | 78.12±0.62 | 65.91±0.89 | 12.45±0.64 | 78.12±0.30 | 66.19±0.09 | 11.56±0.42 | 78.12±0.48 |
| | HCP-IMSC | 2 | 85.63±0.00 | 26.25±0.00 | 85.63±0.00 | 77.68±0.15 | 19.01±0.16 | 78.12±0.00 | 73.45±0.00 | 17.51±0.00 | 78.12±0.00 |
| | APADC | 2 | 72.06±0.57 | 0.08±0.01 | 78.12±0.29 | 66.06±0.53 | 0.06±0.01 | 78.12±0.04 | 60.32±0.90 | 0.06±0.02 | 78.12±0.93 |
| | HCLS-CGL | 2 | 77.64±0.00 | 0.42±0.00 | 78.12±0.00 | 77.74±0.00 | 0.34±0.00 | 78.12±0.00 | 77.64±0.00 | 0.42±0.00 | 78.12±0.00 |
| | Ours-1 | 0 | 86.20±0.00 | 32.61±0.00 | 86.20±0.00 | 85.35±0.00 | 31.95±0.00 | 85.35±0.00 | 73.83±0.00 | 4.11±0.00 | 78.12±0.00 |
| | Ours-2 | 0 | 86.20±0.00 | 32.61±0.00 | 86.20±0.00 | 85.35±0.00 | 31.95±0.00 | 85.35±0.00 | 73.83±0.00 | 4.11±0.00 | 78.12±0.00 |
| Wikifea | IMSC-AGL | 3 | 16.73±0.57 | 2.47±0.34 | 18.83±0.32 | 16.82±0.72 | 3.07±0.41 | 19.94±0.64 | 17.42±0.51 | 4.10±0.32 | 20.75±0.36 |
| | AWP | 0 | 48.01±0.00 | 44.95±0.00 | 57.92±0.00 | 45.43±0.00 | 38.67±0.00 | 50.47±0.00 | 40.86±0.00 | 30.66±0.00 | 44.51±0.00 |
| | APMC | 2 | 41.13±0.04 | 32.49±0.03 | 48.77±0.03 | 40.89±0.07 | 29.34±0.11 | 45.50±0.07 | 37.25±0.05 | 27.61±0.05 | 44.07±0.05 |
| | IMG | 3 | 12.70±0.27 | 0.65±0.04 | 16.09±0.07 | 12.95±0.19 | 0.63±0.05 | 15.94±0.08 | 13.34±0.25 | 0.56±0.04 | 16.08±0.09 |
| | TMBSD | 2 | 42.45±0.06 | 43.99±0.06 | 41.00±0.06 | 37.30±0.11 | 38.65±0.11 | 36.05±0.11 | 35.25±0.07 | 35.25±0.07 | 34.12±0.07 |
| | IKMKC | 2 | 53.26±0.04 | 45.25±0.02 | 55.33±0.03 | 48.83±0.02 | 40.38±0.03 | 50.04±0.04 | 44.08±0.07 | 35.02±0.05 | 45.63±0.07 |
| | IMVTSC-MVI | 3 | 48.11±3.90 | 45.35±2.42 | 53.79±3.46 | 45.10±3.42 | 41.23±2.48 | 50.63±3.18 | 44.08±2.96 | 32.82±1.28 | 46.62±1.86 |
| | CPM-Nets | 1 | 37.46±0.15 | 24.62±0.88 | 38.36±0.90 | 25.63±0.45 | 17.86±0.08 | 27.98±0.15 | 28.07±0.81 | 19.39±0.14 | 31.26±0.18 |
| | LSIMVC | 4 | 32.94±0.00 | 30.07±0.00 | 33.91±0.00 | 30.84±0.00 | 27.90±0.00 | 31.89±0.00 | 28.55±0.02 | 25.07±0.01 | 29.45±0.00 |
| | GSRIMC | 3 | 37.84±3.60 | 27.55±3.30 | 42.48±3.62 | 38.22±3.05 | 28.35±2.32 | 42.73±2.65 | 38.83±2.46 | 26.09±1.66 | 43.11±1.92 |
| | COMPLETER | 3 | 48.29±0.80 | 40.69±0.82 | 51.45±0.14 | 44.12±0.09 | 34.32±0.53 | 49.39±0.80 | 45.43±0.60 | 31.28±0.86 | 47.32±0.10 |
| | TCIMC | 3 | 15.91±0.00 | 0.67±0.00 | 16.05±0.00 | 15.81±0.02 | 0.61±0.02 | 15.99±0.01 | 15.84±0.01 | 0.70±0.02 | 15.98±0.01 |
| | LRGR-IMVC | 2 | 48.01±0.01 | 40.08±0.01 | 55.27±0.00 | 46.52±0.05 | 35.00±0.04 | 51.48±0.02 | 37.92±0.01 | 27.27±0.01 | 43.93±0.01 |
| | BGIMVSC | 2 | 18.26±0.56 | 6.28±1.00 | 18.82±0.53 | 17.58±0.28 | 4.50±0.53 | 18.01±0.28 | 17.48±0.62 | 4.64±0.92 | 17.98±0.60 |
| | NGSP-CGL | 3 | 50.01±1.29 | 42.14±0.52 | 52.70±0.95 | 41.19±1.95 | 34.78±0.41 | 45.64±1.00 | 34.73±1.26 | 28.03±1.08 | 39.80±1.17 |
| | PIMVC | 2 | 49.86±0.13 | 38.34±0.17 | 54.50±0.12 | 46.85±0.23 | 34.62±0.05 | 50.84±0.07 | 40.09±0.27 | 29.69±0.24 | 43.64±0.26 |
| | ProImp | 2 | 51.92±0.74 | 43.62±0.42 | 57.93±0.56 | 46.28±0.64 | 35.88±0.75 | 51.43±0.02 | 45.16±0.85 | 30.62±0.46 | 48.40±0.95 |
| | HCP-IMSC | 2 | 37.81±0.25 | 23.78±0.11 | 42.92±0.16 | 35.06±0.04 | 19.29±0.04 | 39.64±0.05 | 37.48±0.03 | 22.06±0.04 | 42.91±0.05 |
| | APADC | 2 | 42.15±0.62 | 33.89±0.72 | 44.17±0.49 | 32.07±0.82 | 26.25±0.64 | 36.86±0.01 | 26.60±0.97 | 21.81±0.71 | 31.06±0.11 |
| | HCLS-CGL | 2 | 53.44±0.92 | 44.13±0.53 | 58.20±0.41 | 46.94±1.00 | 37.93±0.22 | 50.79±0.43 | 41.88±0.50 | 29.60±0.54 | 46.19±0.75 |
| | Ours-1 | 0 | 56.28±0.00 | 47.90±0.00 | 58.86±0.00 | 48.95±0.00 | 41.67±0.00 | 51.74±0.00 | 44.17±0.00 | 35.63±0.00 | 47.07±0.00 |
| | Ours-2 | 0 | 56.28±0.00 | 47.90±0.00 | 58.86±0.00 | 48.95±0.00 | 41.67±0.00 | 51.74±0.00 | 44.17±0.00 | 35.63±0.00 | 47.07±0.00 |
| AWA10 | IMSC-AGL | 3 | 13.81±0.33 | 1.02±0.12 | 20.10±0.03 | 13.65±0.35 | 0.82±0.08 | 20.09±0.00 | 12.99±0.43 | 0.74±0.06 | 20.13±0.07 |
| | AWP | 0 | 21.36±0.00 | 9.75±0.00 | 22.65±0.00 | 22.03±0.00 | 9.90±0.00 | 22.96±0.00 | 21.95±0.00 | 9.34±0.00 | 22.67±0.00 |
| | APMC | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | IMG | 3 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | TMBSD | 2 | 15.18±0.06 | 11.51±0.07 | 14.05±0.05 | 15.26±0.05 | 11.69±0.05 | 14.06±0.05 | 14.47±0.08 | 8.84±0.09 | 13.32±0.08 |
| | IKMKC | 2 | 21.90±0.17 | 7.50±0.07 | 23.58±0.09 | 19.21±0.17 | 6.11±0.19 | 21.88±0.09 | 18.39±0.19 | 5.71±0.27 | 21.75±0.13 |
| | IMVTSC-MVI | 3 | 26.58±1.13 | 12.12±0.23 | 29.97±0.74 | 23.06±0.50 | 11.04±0.25 | 27.30±0.55 | 22.37±0.73 | 9.03±0.37 | 25.22±0.46 |
| | CPM-Nets | 1 | 22.16±0.27 | 9.82±0.07 | 25.07±0.66 | 21.76±0.30 | 9.06±0.40 | 24.60±0.37 | 20.74±0.38 | 9.88±0.17 | 25.56±0.96 |
| | LSIMVC | 4 | 23.00±0.24 | 7.90±0.12 | 24.29±0.25 | 21.87±0.29 | 6.68±0.09 | 23.71±0.23 | 21.09±0.20 | 6.33±0.16 | 23.44±0.18 |
| | GSRIMC | 3 | 24.37±1.05 | 10.97±0.93 | 27.75±0.89 | 22.65±1.21 | 10.05±0.86 | 26.66±0.81 | 22.05±1.02 | 9.12±0.25 | 26.01±0.82 |
| | COMPLETER | 3 | 20.24±0.74 | 5.64±0.21 | 22.19±0.30 | 18.99±0.08 | 5.79±0.99 | 22.27±0.60 | 22.15±0.01 | 5.85±0.32 | 22.91±0.75 |
| | TCIMC | 3 | 22.18±0.07 | 1.24±0.17 | 22.10±0.09 | 21.86±0.06 | 1.21±0.21 | 21.59±0.09 | 20.46±0.08 | 1.10±0.20 | 20.59±0.10 |
| | LRGR-IMVC | 2 | 24.64±0.13 | 12.36±0.05 | 28.27±0.15 | 24.47±0.05 | 12.05±0.04 | 27.32±0.03 | 22.48±0.29 | 9.07±0.11 | 25.41±0.14 |
| | BGIMVSC | 2 | 20.22±0.01 | 0.42±0.02 | 20.31±0.01 | 20.19±0.02 | 0.39±0.03 | 20.24±0.02 | 20.20±0.01 | 0.36±0.01 | 20.23±0.01 |
| | NGSP-CGL | 3 | 21.26±0.63 | 7.21±0.82 | 23.73±0.59 | 19.32±1.04 | 5.62±0.82 | 22.97±0.74 | 19.19±1.06 | 5.27±0.71 | 23.46±0.63 |
| | PIMVC | 2 | 24.96±0.12 | 11.90±0.06 | 28.32±0.13 | 25.44±0.26 | 11.91±0.14 | 26.62±0.37 | 24.28±0.05 | 8.72±0.03 | 24.52±0.05 |
| | ProImp | 2 | 22.82±0.57 | 9.91±0.37 | 27.00±0.98 | 22.49±0.28 | 10.06±0.04 | 27.25±0.89 | 20.48±0.63 | 8.56±0.69 | 25.10±0.07 |
| | HCP-IMSC | 2 | 24.44±0.04 | 9.98±0.03 | 27.81±0.05 | 23.84±0.15 | 10.13±0.16 | 27.44±0.11 | 24.47±0.27 | 8.99±0.06 | 25.16±0.13 |
| | APADC | 2 | 15.43±0.26 | 2.24±0.16 | 20.92±0.95 | 15.57±0.93 | 2.05±0.54 | 20.57±0.33 | 15.45±0.52 | 1.76±0.18 | 20.31±0.25 |
| | HCLS-CGL | 2 | 20.80±0.47 | 7.92±0.07 | 23.40±0.12 | 21.14±0.23 | 7.71±0.02 | 23.67±0.02 | 22.32±0.12 | 7.79±0.05 | 24.15±0.07 |
| | Ours-1 | 0 | 28.88±0.00 | 13.25±0.00 | 30.60±0.00 | 26.37±0.00 | 12.31±0.00 | 27.76±0.00 | 24.63±0.00 | 9.41±0.00 | 26.16±0.00 |
| | Ours-2 | 0 | 28.88±0.00 | 13.25±0.00 | 30.60±0.00 | 26.37±0.00 | 12.31±0.00 | 27.76±0.00 | 24.63±0.00 | 9.41±0.00 | 26.16±0.00 |

The optimal solution $\mathbf{E}_m$ is

$$\mathbf{E}_m = \mathbf{D}_m \mathbf{W}_m \mathbf{W}_m^\top \mathbf{L}^\top \mathbf{O}^\top \left(\mathbf{O}\mathbf{A}_m\mathbf{O}^\top\right)^{-1}. \tag{21}$$

**Algorithm** 3 gives the whole procedure for solving Eq. (4).

For the computational complexity of ToRES, we have

**Theorem 3.5.** *ToRES's computational complexity is* $\mathcal{O}(n)$.

In virtue of **Theorem 3.5**, ToRES can efficiently tackle large-scale IMVC tasks.

# 4. Experiments and Analysis

## 4.1. Benchmark Datasets and Competitors

Six widely-used datasets are adopted in experiments, including small-scale datasets: **Webkb**, **Wikifea**; middle-

*Table 2.* Average Clustering Results on Datasets SUNRGB-D, AwAfea and EMNIST.

| Dataset | Methods | NoHp | 20% | | | 40% | | | 60% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | ACC | NMI | Purity | ACC | NMI | Purity | ACC | NMI | Purity |
| SUNRGB-D | IMSC-AGL | 3 | 7.13±0.30 | 5.95±0.16 | 17.20±0.38 | 6.69±0.18 | 5.85±0.14 | 16.19±0.28 | 6.76±0.21 | 5.63±0.15 | 16.33±0.25 |
| | AWP | 0 | 17.61±0.00 | 26.08±0.00 | 36.22±0.00 | 17.46±0.00 | 22.08±0.00 | 34.25±0.00 | 16.94±0.00 | 20.77±0.00 | 34.15±0.00 |
| | APMC | 2 | 20.79±0.75 | 25.02±0.33 | 35.32±0.71 | 17.27±0.51 | 22.94±0.26 | 33.27±0.50 | 16.41±0.45 | 19.83±0.31 | 32.65±0.43 |
| | IMG | 3 | 6.77±0.15 | 5.76±0.12 | 16.65±0.17 | 5.97±0.13 | 5.35±0.10 | 15.62±0.21 | 6.11±0.16 | 6.21±0.13 | 16.54±0.19 |
| | TMBSD | 2 | 11.54±0.13 | 22.22±0.24 | 7.79±0.10 | 11.38±0.25 | 21.76±0.45 | 7.70±0.17 | 10.36±0.33 | 18.98±0.63 | 7.13±0.23 |
| | IKMKC | 2 | 19.05±0.66 | 23.26±0.30 | 36.04±0.44 | 17.42±0.44 | 20.84±0.24 | 33.19±0.69 | 15.97±0.39 | 19.10±0.27 | 31.17±0.65 |
| | IMVTSC-MVI | 3 | 16.61±0.56 | 20.27±0.18 | 32.57±0.43 | 15.65±0.52 | 16.94±0.15 | 29.14±0.42 | 14.11±0.55 | 13.75±0.12 | 25.53±0.25 |
| | CPM-Nets | 1 | 14.04±0.09 | 17.71±0.18 | 28.52±0.90 | 13.87±0.62 | 16.76±0.86 | 26.98±0.67 | 12.55±0.30 | 14.87±0.45 | 26.02±0.95 |
| | LSIMVC | 4 | 11.65±0.17 | 18.16±0.13 | 29.67±0.34 | 12.29±0.22 | 18.23±0.08 | 29.66±0.22 | 13.13±0.25 | 18.36±0.16 | 29.96±0.29 |
| | GSRIMC | 3 | 17.64±0.49 | 25.62±0.25 | 36.88±0.32 | 17.03±0.40 | 24.26±0.28 | 35.02±0.38 | 15.81±0.37 | 20.24±0.21 | 34.08±0.48 |
| | COMPLETER | 3 | 17.84±0.89 | 16.80±0.80 | 25.50±0.04 | 18.91±0.01 | 15.80±0.11 | 25.02±0.13 | 19.34±0.51 | 15.01±0.34 | 23.70±0.44 |
| | TCIMC | 3 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | LRGR-IMVC | 2 | 18.61±0.35 | 25.01±0.28 | 39.81±0.38 | 17.11±0.22 | 23.76±0.17 | 35.11±0.28 | 16.04±0.29 | 20.61±0.23 | 34.14±0.23 |
| | BGIMVSC | 2 | 14.33±1.52 | 7.34±2.25 | 15.04±1.67 | 11.09±0.31 | 3.05±0.55 | 11.75±0.32 | 10.95±0.10 | 2.94±0.14 | 11.57±0.08 |
| | NGSP-CGL | 3 | 18.36±0.67 | 23.43±0.37 | 35.87±0.59 | 15.68±0.51 | 20.89±0.27 | 32.20±0.38 | 15.19±0.46 | 19.53±0.28 | 30.09±0.67 |
| | PIMVC | 2 | 16.09±0.41 | 24.47±0.21 | 35.92±0.31 | 14.96±0.21 | 22.82±0.15 | 34.55±0.25 | 14.95±0.25 | 20.16±0.22 | 33.42±0.48 |
| | ProImp | 2 | 13.61±0.66 | 19.88±0.93 | 30.90±0.55 | 12.40±0.76 | 17.86±0.92 | 28.91±0.82 | 11.67±0.00 | 16.92±0.01 | 28.37±0.95 |
| | HCP-IMSC | 2 | 17.50±0.40 | 24.24±0.27 | 36.76±0.28 | 17.61±0.39 | 23.40±0.21 | 34.42±0.45 | 16.27±0.32 | 20.08±0.33 | 35.82±0.46 |
| | APADC | 2 | 9.44±0.35 | 8.77±0.09 | 19.99±0.90 | 9.46±0.31 | 8.31±0.89 | 19.22±0.17 | 8.42±0.54 | 7.88±0.33 | 18.72±0.09 |
| | HCLS-CGL | 2 | 20.20±0.45 | 24.29±0.18 | 37.01±0.27 | 18.62±0.46 | 22.24±0.19 | 35.16±0.27 | 17.36±0.49 | 21.64±0.15 | 34.56±0.31 |
| | Ours-1 | 0 | 20.93±0.00 | 25.73±0.00 | 37.16±0.00 | 19.82±0.00 | 23.87±0.00 | 35.94±0.00 | 19.75±0.00 | 20.98±0.00 | 32.31±0.00 |
| | Ours-2 | 0 | 20.93±0.00 | 25.73±0.00 | 37.16±0.00 | 19.82±0.00 | 23.87±0.00 | 35.94±0.00 | 19.75±0.00 | 20.98±0.00 | 32.31±0.00 |
| AwAfea | IMSC-AGL | 3 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | AWP | 0 | 8.25±0.00 | 9.37±0.00 | 9.29±0.00 | 8.71±0.00 | 9.14±0.00 | 10.00±0.00 | 8.25±0.00 | 8.81±0.00 | 9.15±0.00 |
| | APMC | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | IMG | 3 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | TMBSD | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | IKMKC | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | IMVTSC-MVI | 3 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | CPM-Nets | 1 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | LSIMVC | 4 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | GSRIMC | 3 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | COMPLETER | 3 | 7.00±0.94 | 7.62±0.31 | 7.73±0.11 | 6.63±0.22 | 7.41±0.49 | 7.71±0.38 | 6.93±0.96 | 7.95±0.90 | 7.99±0.49 |
| | TCIMC | 3 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | LRGR-IMVC | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | BGIMVSC | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | NGSP-CGL | 3 | 6.46±0.19 | 6.00±0.31 | 7.24±0.26 | 5.90±0.19 | 5.22±0.31 | 6.72±0.24 | 5.70±0.17 | 5.06±0.27 | 6.65±0.19 |
| | PIMVC | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | ProImp | 2 | 7.73±0.09 | 9.67±0.23 | 9.86±0.55 | 7.40±0.27 | 9.04±0.04 | 9.54±0.16 | 7.09±0.68 | 8.35±0.79 | 9.07±0.06 |
| | HCP-IMSC | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | APADC | 2 | 4.92±0.00 | 3.05±0.76 | 5.82±0.09 | 4.72±0.21 | 3.02±0.53 | 5.91±0.43 | 4.52±0.58 | 2.81±0.46 | 5.54±0.45 |
| | HCLS-CGL | 2 | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | Ours-1 | 0 | 8.96±0.00 | 11.17±0.00 | 10.29±0.00 | 8.72±0.00 | 10.62±0.00 | 10.43±0.00 | 8.62±0.00 | 10.33±0.00 | 10.21±0.00 |
| | Ours-2 | 0 | 8.96±0.00 | 11.17±0.00 | 10.29±0.00 | 8.72±0.00 | 10.62±0.00 | 10.43±0.00 | 8.62±0.00 | 10.33±0.00 | 10.21±0.00 |
| EMNIST | All Compared Algorithms | | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| | Ours-1 | 0 | 47.18±0.00 | 44.27±0.00 | 48.27±0.00 | 43.23±0.00 | 44.57±0.00 | 44.29±0.00 | 45.22±0.00 | 45.45±0.00 | 48.52±0.00 |
| | Ours-2 | 0 | 47.18±0.00 | 44.27±0.00 | 48.27±0.00 | 43.23±0.00 | 44.57±0.00 | 44.29±0.00 | 45.22±0.00 | 45.45±0.00 | 48.52±0.00 |

scale datasets: **AWA10**, **SUNRGB-D**; large-scale datasets: **AwAfea**, **EMNIST**. Table 3 describes their details.

*Table 3.* Dataset Details.

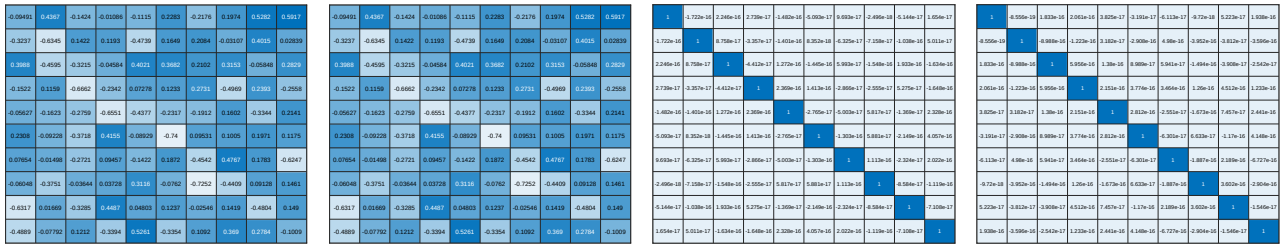
| Dataset | Samples | Clusters | Views | Feature Dimension |
|---|---|---|---|---|
| Webkb | 1051 | 2 | 2 | 334,2949 |
| Wikifea | 2866 | 10 | 2 | 128,10 |
| AWA10 | 5814 | 10 | 6 | 2688,2000,252,2000,2000,2000 |
| SUNRGB-D | 10335 | 45 | 2 | 4096,4096 |
| AwAfea | 30475 | 50 | 6 | 2688,2000,252,2000,2000,2000 |
| EMNIST | 280000 | 10 | 4 | 944,512,576,640 |

The following 20 classical IMVC algorithms are adopted as the baselines to highlight the effectiveness of ToRES: **IMSC-AGL** (Wen et al., 2020), **AWP** (Nie et al., 2018), **APMC** (Guo & Ye, 2019), **IMG** (Zhao et al., 2016), **TMBSD** (Li et al., 2021c), **IKMKC** (Liu et al., 2020), **IMVTSC-MVI** (Wen et al., 2021b), **CPM-Nets** (Zhang et al., 2019), **LSIMVC** (Liu et al., 2022a), **GSRIMC** (Li et al., 2022a), **COMPLETER** (Lin et al., 2021), **TCIMC** (Xia et al., 2022a), **LRGR-IMVC** (Cui et al., 2022), **BGIMVSC** (Sun et al., 2023), **NGSP-CGL** (Wong et al., 2023), **PIMVC** (Deng et al., 2023), **ProImp** (Li et al., 2023a), **HCP-IMSC** (Li et al., 2022b), **APADC** (Xu et al., 2023), **HCLS-CGL** (Wen et al., 2023).

### 4.2. Observations and Analysis

We summary the clustering results under the percentage of incomplete data (PID) = 20%, 40% and 60% in Table 1 and 2, where "Ours-1" and "Ours-2" represent the results obtained based on Algorithm 1 and Algorithm 2, respectively. "NoHp" represents the number of hyper-parameters. From these two tables, we can draw the following observations:

**1)** Our variance is 0 on all datasets. This is because we learn to directly produce the labels rather than first forming spectral embedding and then performing clustering operation on it. This not only ensures clustering process and optimization procedure to interact and facilitate mutually, but also makes



(a) $\mathbf{O}_{Algo1}$ (b) $\mathbf{O}_{Algo2}$ (c) $\mathbf{O}_{Algo1}^{\top}\mathbf{O}_{Algo1}$ (d) $\mathbf{O}_{Algo2}^{\top}\mathbf{O}_{Algo2}$

*Figure 2.* Visualization of $\mathbf{O}$ learned on AWA10 under PID = 40 %. $\mathbf{O}_{Algo1}$ is the $\mathbf{O}$ learned by Algorithm 1 while $\mathbf{O}_{Algo2}$ is that learned by Algorithm 2. $\mathbf{O}_{Algo1}^{\top}\mathbf{O}_{Algo1}$ is the matrix product result of $\mathbf{O}_{Algo1}^{\top}$ and $\mathbf{O}_{Algo1}$. Similar meaning is for $\mathbf{O}_{Algo2}^{\top}\mathbf{O}_{Algo2}$.

the clustering labels directly dependent on the original data, thereby guaranteeing the results to be enhanced and stable.

**2)** Ours-1 has the same clustering results as Ours-2. This indicates that the solutions acquired by Algorithm 1 and Algorithm 2 are equivalent. To verify this point, we visualize the learned solutions. As seen from Fig. 2 (a) and (b), the solution $\mathbf{O}$ learned by Algorithm 1 is indeed equivalent to that learned by Algorithm 2. Also, Fig. 2 (c) and (d) suggest that these learned solutions satisfy the constraint condition.

**3)** APMC and IMG are incapable of tackling AWA10 since they only work with two or three views, which largely impedes their practicality and also is not conductive to extracting sufficient representation for superior clustering results. For example, on Wikifea and SUNRGB-D, they are clearly inferior against ours. By comparison, ToRES does not encounter this limitation, and is applicable to multiple views.

**4)** Compared to AWP, IMG, IKMKC, LSIMVC, BGIMVSC, HCP-IMSC, PIMVC and etc, which handle IMVC tasks via subspace, NMF or kernel scheme, ToRES based on prototype makes preferable results in most cases. For example, on Webkb with PID = 20%, we are higher than them by 13.03%, 35.39%, 36.06%, 18.84%, 0.57%, 8.37%, 4.09% in ACC. This indicates that our prototype strategy is functional.

**5)** In comparison with CPM-Nets, COMPLETER, ProImp and APADC that handle IMVC problems via genious neural networks, our ToRES consistently exceeds all of them on Webkb, Wikefea, AWA10, SUNRGBD and AwAfea under PID = 20% and 40%. Under PID=60%, ToRES still can generate comparable results. This well suggests that even facing deep methods, ToRES is still with strong competitiveness.

**6)** ToRES produces some sub-optimal results under PID = 60%, for example on Webkb. One reason could be that the learned prototypes are misaligned due to the absence of cluster information, and higher PID further exacerbates this phenomenon. Another reason is that substantial incomplete data leads to cluster imbalance and deteriorates the quality of learned prototypes, harming the clustering performance.

**7)** IMVTS-MVI receives several slightly better results than ours, the reason of which could be that it successfully embeds a consensus low-rank tensor constraint into representation learning to maintain the semantic consistency, and eliminates the gab between manifold space and feature space by jointly running similarity graph recovery and incomplete view inferring to deeply mine features hidden within data.

**8)** On SUNRGB-D with PID = 20%, AWP outperforms us by 0.35% in NMI. Possible reasons are that AWP employs a unified indicator to alleviate the disagreements among views instead of separately evaluating each view, and seamlessly integrates inter-view affinity graph with different scales by introducing weighting strategy to adaptively adjust view capacity, thereby capturing superior clustering representation.

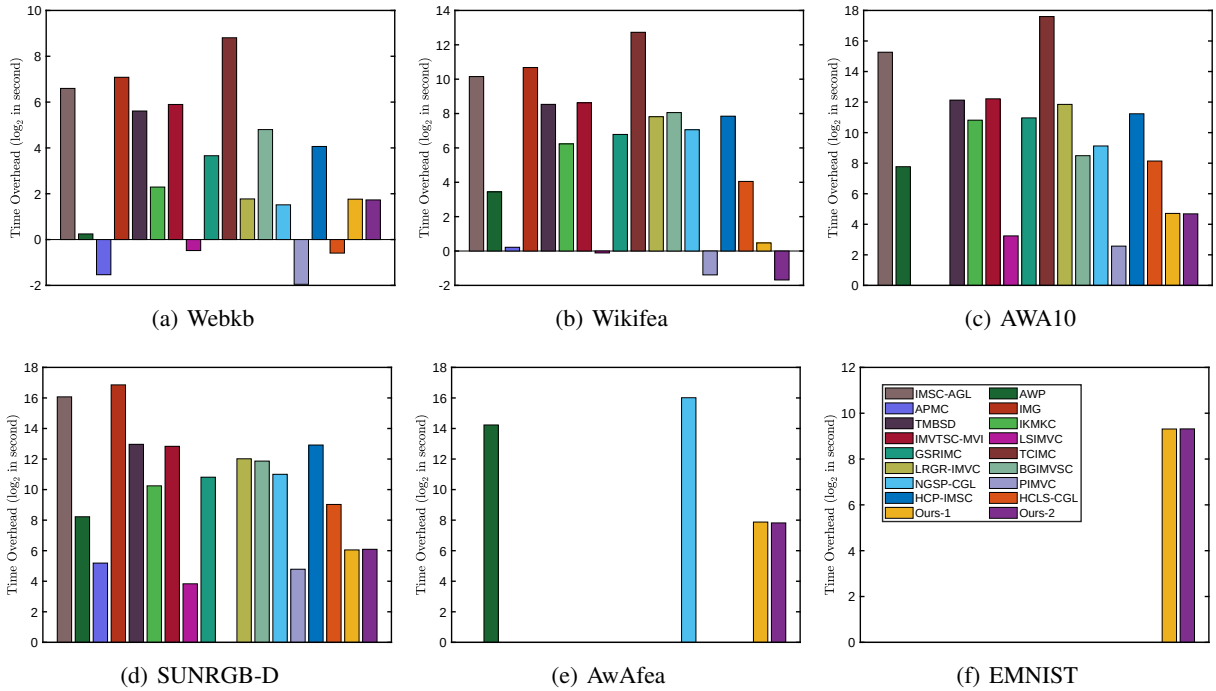
**9)** For GSRIMC, it is higher than ours by 0.39% and 1.77% in NMI on SUNRGB-D under PID = 40% and 60% respectively, possibly because GSRIMC reduces the intra-cluster density relation by separating the biased errors caused by incomplete data from similarity graph, and refines the graph structure of the boundary instances by explicitly preserving the local structure information using joint-graph learning.

**10)** LRGR-IMVC makes a 2.65% improvement in terms of Purity than ours on SUNRGB-D under PID = 20%. Main reasons could be that LRGR-IMVC maintains the semantic correlation by exploring the latent relationship between views using a graph rank constraint, and restores the incomplete affinity relationship adaptively by integrating the global structure and underlying information between views.

### 4.3. Time Overhead

As shown in **Theorem 3.5**, ToRES owns linear computing cost. To further highlight ToRES's time-friendly characteristic, we count the running time in Fig. 3. We can observe

**1)** TCIMC encounters the slowest running speed. Particularly, on AWA10 that contains only 5814 samples, it consumes 198282.32 seconds. This is mainly due to the introduction of tensor constraints, which typically bring about




(a) Webkb (b) Wikifea (c) AWA10

(d) SUNRGB-D (e) AwAfea (f) EMNIST

*Figure 3.* Time Overhead Comparison.

higher computing cost. Methods such as TMBSD, IMVTSC-MVI, GSRIMC, etc, similarly suffer from such restrictions.

**2)** IMSC-AGL, IMG, BGIMVSC, AWP, HCP-IMSC, etc, require more time than us. The reasons are that these subspace based methods typically need to construct self-expression relation. In contrast, ToRES successfully builds prototype-sample affinity with small size, and also does not involve spectral embedding, thus consuming relatively less time.

**3)** IKMKC, NGSP-CGL and HCLS-CGL are slower than us, which is mainly because IKMKC requires constructing the kernel with full size, NGSP-CGL needs to count the nearest neighbor sets of all sample pairs, and HCLS-CGL introduces additional confidence graphs. This results in them experiencing a at least quadratic computational complexity.

**4)** APMC, LSIMVC and PIMVC take slightly less time than us because of the non-iterative scheme and diagonal element reciprocal strategy. Nevertheless, due to APMC ignoring the view-specific truncated similarities, LSIMVC separately exploiting the geometric structure and representation, and PIMVC only employing the simplest linear relation to fit the space mapping, they generally generate inferior results.

**5)** ToRES can work properly on all of these datasets, which include small-scale, middle-scale and large-scale datasets. Especially, on EMNIST with 280000 samples, it only consumes 635.10 seconds. This illustrates that ToRES is time-friendly and can efficiently tackle datasets with diverse sizes.

### 4.4. Memory Overhead

We also measure the memory overhead to highlight ToRES's memory-friendly characteristic. As reported in Table 4, ToRES owns the lowest memory requirements in most cases.

*Table 4.* Memory Overhead Comparison (GB).

| Methods | Webkb | Wikifea | AWA10 | SUNRGB-D | AwAfea | EMNIST |
|---|---|---|---|---|---|---|
| IMSC-AGL | 0.29 | 1.59 | 13.86 | 18.70 | \ | \ |
| AWP | 0.21 | 1.30 | 11.09 | 18.18 | 94.46 | \ |
| APMC | **0.09** | 0.23 | \ | 3.75 | \ | \ |
| IMG | 0.16 | 0.69 | \ | 10.48 | \ | \ |
| TMBSD | 0.33 | 1.88 | 22.77 | 26.82 | \ | \ |
| IKMKC | 0.17 | 1.49 | 10.05 | 20.71 | \ | \ |
| IMVTSC-MVI | 0.27 | 1.57 | 20.78 | 23.04 | \ | \ |
| LSIMVC | 0.13 | 0.48 | 5.34 | 6.60 | \ | \ |
| GSRIMC | 0.30 | 2.52 | 29.81 | 33.91 | \ | \ |
| TCIMC | 0.49 | 2.92 | 33.18 | \ | \ | \ |
| LRGR-IMVC | 0.20 | 1.09 | 11.34 | 15.92 | \ | \ |
| BGIMVSC | 0.17 | 0.92 | 8.88 | 20.13 | \ | \ |
| NGSP-CGL | 0.26 | 1.95 | 14.24 | 27.58 | 126.71 | \ |
| PIMVC | 0.44 | 0.62 | 4.19 | 7.40 | \ | \ |
| HCP-IMSC | 0.35 | 1.61 | 19.59 | 24.03 | \ | \ |
| HCLS-CGL | 0.20 | 2.00 | 13.92 | 27.22 | \ | \ |
| Ours | 0.22 | **0.20** | **2.36** | **3.68** | **11.58** | **26.34** |

### 4.5. Ablation Studies

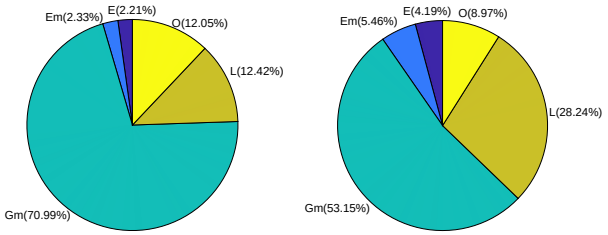
To investigate the effectiveness of each type of prototypes, we organize some ablation studies, and present the results in Table 5, where "CVP " and "VWP" represent the results acquired based on cross-view prototypes and view-wise prototypes respectively. As seen, we can integrate these two parts' information to generate better results. More ablation studies are presented in Section K of the Appendix.

*Table 5.* Ablation Study for View-wise and Cross-view Prototypes

| Dataset | Ablation Study | 20% | | | 40% | | | 60% | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ACC | NMI | Purity | ACC | NMI | Purity | ACC | NMI | Purity |
| Webkb | CVP | 64.99 | 1.90 | 78.12 | 64.99 | 1.66 | 78.12 | 69.17 | 1.81 | 78.12 |
| | VWP | 51.19 | 3.50 | 78.12 | 51.95 | 2.72 | 78.12 | 56.90 | 2.94 | 78.12 |
| | Ours | **86.20** | **32.61** | **86.20** | **85.35** | **31.95** | **85.35** | **73.83** | **4.11** | **78.12** |
| Wikifea | CVP | 54.47 | 46.91 | 57.78 | 47.87 | **41.79** | 50.45 | 42.25 | 34.82 | 45.88 |
| | VWP | 52.51 | 46.18 | 57.50 | 48.89 | 40.09 | 51.31 | **44.80** | 33.04 | 46.19 |
| | Ours | **56.28** | **47.90** | **58.86** | **48.95** | 41.67 | **51.74** | 44.17 | **35.63** | **47.07** |
| AWA10 | CVP | 16.51 | 2.84 | 21.53 | 15.63 | 2.59 | 21.04 | 14.55 | 2.00 | 20.55 |
| | VWP | 16.70 | 1.84 | 20.11 | 17.05 | 1.52 | 20.09 | 16.46 | 1.04 | 20.14 |
| | Ours | **28.88** | **13.25** | **30.60** | **26.37** | **12.31** | **27.76** | **24.63** | **9.41** | **26.16** |
| SUNRGB-D | CVP | 15.52 | 19.00 | 31.68 | 13.72 | 17.51 | 30.27 | 12.41 | 15.67 | 28.15 |
| | VWP | 15.40 | 5.98 | 16.23 | 13.50 | 4.01 | 14.11 | 12.52 | 3.57 | 13.11 |
| | Ours | **20.93** | **25.73** | **37.16** | **19.82** | **23.87** | **35.94** | **19.75** | **20.98** | **32.31** |
| AwAfea | CVP | 6.68 | 6.88 | 7.99 | 6.41 | 6.27 | 7.98 | 5.80 | 5.53 | 7.30 |
| | VWP | 4.23 | 1.44 | 4.34 | 4.38 | 1.42 | 4.63 | 4.20 | 1.36 | 4.32 |
| | Ours | **8.96** | **11.17** | **10.29** | **8.72** | **10.62** | **10.43** | **8.62** | **10.33** | **10.21** |
| EMNIST | CVP | 37.15 | 24.68 | 39.40 | 31.30 | 20.86 | 33.81 | 30.66 | 18.16 | 34.17 |
| | VWP | 13.31 | 2.76 | 13.31 | 16.91 | 5.92 | 16.91 | 16.50 | 5.36 | 16.60 |
| | Ours | **47.18** | **44.27** | **48.27** | **43.23** | **44.57** | **44.29** | **45.22** | **45.45** | **48.52** |

## 5. Limitation and Future Work

In the paper we equally treat cross-view prototypes and view-wise prototypes. Devising a weighting scheme for them could be more reasonable, as views generally have different levels of importance. Besides, we count the time overhead proportion of optimization variables. As seen in Fig. 4, $\mathbf{G}_m$ takes the most time, which is mainly because it requires performing SVD on the matrix with size $d_m \times d_m$. Thus, providing more innovative solutions for $\mathbf{G}_m$ in the future could further accelerate the running speed of ToRES.




(a) AwAfea (b) EMNIST

*Figure 4.* Time Overhead Proportion of $\mathbf{E}_m$, $\mathbf{E}$, $\mathbf{O}$, $\mathbf{L}$ and $\mathbf{G}_m$.

## 6. Conclusion

In this work, we devise an IMVC algorithm with resource-friendly and extensible as well as stable properties, named ToRES. Instead of self-expression affinity, which not only contains blanks but also requires intensive resource consumption, ToRES successfully constructs prototype-sample affinity with low complexity. Specially, it introduces two types of prototypes to jointly exploit preferable representation. Owing to not involving hyper-parameters, it is easily extended to other scenes. Also, it directly learns discrete cluster labels from incomplete data, effectively preserving original information and ensuring result stability. Two equivalent solutions are developed for the resulting problem. Experimental results well demonstrate ToRES's advantages.

## Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

## Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (No. 2022ZD0209103); in part by the National Natural Science Foundation of China (No. 62325604 and 62276271).

## References

Chen, Y., Xiao, X., Peng, C., Lu, G., and Zhou, Y. Low-rank tensor graph learning for multi-view subspace clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(1):92–104, 2021.

Cui, J., Fu, Y., Huang, C., and Wen, J. Low-rank graph completion-based incomplete multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

Deng, S., Wen, J., Liu, C., Yan, K., Xu, G., and Xu, Y. Projective incomplete multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

Fu, S., Peng, Q., He, Y., et al. Towards unsupervised graph completion learning on graphs with features and structure missing. In *2023 IEEE International Conference on Data Mining*, pp. 1019–1024. IEEE, 2023.

Fu, S., Peng, Q., He, Y., et al. Multilevel contrastive graph masked autoencoders for unsupervised graph-structure learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

Guo, J. and Ye, J. Anchors bring ease: An embarrassingly simple approach to partial multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 118–125, 2019.

Li, H., Li, Y., Yang, M., Hu, P., Peng, D., and Peng, X. Incomplete multi-view clustering via prototype-based imputation. In *International Joint Conference on Artificial Intelligence*, 2023a.

Li, J., Gao, Q., Wang, Q., Yang, M., and Xia, W. Orthogonal non-negative tensor factorization based multi-view clustering. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023b.

Li, L., Wan, Z., and He, H. Incomplete multi-view clustering with joint partition and graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021a.

Li, L., Zhang, J., Wang, S., Liu, X., Li, K., and Li, K. Multi-view bipartite graph clustering with coupled noisy feature filter. *IEEE Transactions on Knowledge and Data Engineering*, 2023c.

Li, L., Pan, Y., Liu, J., Liu, Y., Liu, X., Li, K., Tsang, I. W., and Li, K. Bgae: Auto-encoding multi-view bipartite graph clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

Li, M., Xia, J., Xu, H., Liao, Q., Zhu, X., and Liu, X. Localized incomplete multiple kernel k-means with matrix-induced regularization. *IEEE Transactions on Cybernetics*, 2021b.

Li, R., Zhang, C., Fu, H., Peng, X., Zhou, T., and Hu, Q. Reciprocal multi-layer subspace learning for multi-view clustering. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 8172–8180, 2019.

Li, X.-L., Chen, M.-S., Wang, C.-D., and Lai, J.-H. Refining graph structure for incomplete multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2022a.

Li, Z., Tang, C., Liu, X., Zheng, X., Zhang, W., and Zhu, E. Tensor-based multi-view block-diagonal structure diffusion for clustering incomplete multi-view data. In *IEEE International Conference on Multimedia and Expo*, pp. 1–6. IEEE, 2021c.

Li, Z., Tang, C., Zheng, X., Liu, X., Zhang, W., and Zhu, E. High-order correlation preserved incomplete multi-view subspace clustering. *IEEE Transactions on Image Processing*, 31:2067–2080, 2022b.

Liang, W., Liu, X., Liu, Y., Huang, J.-J., Wang, S., Liu, J., Zhang, Y., Zhu, E., et al. Stability and generalization of kernel clustering: From single kernel to multiple kernel. *Advances in Neural Information Processing Systems*, 35: 33633–33645, 2022a.

Liang, W., Liu, X., Zhou, S., Liu, J., Wang, S., and Zhu, E. Robust graph-based multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7462–7469, 2022b.

Liang, W., Zhou, S., Xiong, J., Liu, X., Wang, S., Zhu, E., Cai, Z., and Xu, X. Multi-view spectral clustering with high-order optimal neighborhood laplacian matrix. *IEEE Transactions on Knowledge & Data Engineering*, 34(07): 3418–3430, 2022c.

Liang, W., Liu, X., Liu, Y., Ma, C., Zhao, Y., Liu, Z., and Zhu, E. Consistency of multiple kernel clustering. In *International Conference on Machine Learning*, pp. 20650–20676. PMLR, 2023.

Liang, W., Tang, C., Liu, X., Liu, Y., Liu, J., Zhu, E., and He, K. On the consistency and large-scale extension of multiple kernel clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Lin, Y., Gou, Y., Liu, Z., Li, B., Lv, J., and Peng, X. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11174–11183, 2021.

Liu, C., Wu, Z., Wen, J., Xu, Y., and Huang, C. Localized sparse incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 2022a.

Liu, X., Zhu, X., Li, M., Wang, L., Zhu, E., Liu, T., Kloft, M., Shen, D., Yin, J., and Gao, W. Multiple kernel k-means with incomplete kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

Liu, X., Liu, L., Liao, Q., Wang, S., Zhang, Y., Tu, W., Tang, C., Liu, J., and Zhu, E. One pass late fusion multi-view clustering. In *International conference on machine learning*, pp. 6850–6859. PMLR, 2021.

Liu, Y., Tu, W., Zhou, S., Liu, X., Song, L., Yang, X., and Zhu, E. Deep graph clustering via dual correlation reduction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7603–7611, 2022b.

Liu, Y., Liang, K., Xia, J., Zhou, S., Yang, X., , Liu, X., and Li, Z. S. Dink-net: Neural clustering on large graphs. In *Proc. of ICML*, 2023.

Ma, H., Wang, S., Zhang, J., Yu, S., Liu, S., Liu, X., and He, K. Symmetric multi-view subspace clustering with automatic neighbor discovery. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

Nie, F., Tian, L., and Li, X. Multiview clustering via adaptively weighted procrustes. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2022–2030, 2018.

Pan, E. and Kang, Z. Multi-view contrastive graph clustering. *Advances in neural information processing systems*, 34:2148–2159, 2021.

Peng, X., Huang, Z., Lv, J., Zhu, H., and Zhou, J. T. Comic: Multi-view clustering without parameter selection. In *International Conference on Machine Learning*, pp. 5092–5101. PMLR, 2019.

Sun, L., Wen, J., Liu, C., Fei, L., and Li, L. Balance guided incomplete multi-view spectral clustering. *Neural Networks*, 2023.

Wan, X., Liu, J., Liang, W., Liu, X., Wen, Y., and Zhu, E. Continual multi-view clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 3676–3684, 2022.

Wan, X., Liu, X., Liu, J., Wang, S., Wen, Y., Liang, W., Zhu, E., Liu, Z., and Zhou, L. Auto-weighted multi-view clustering for large-scale data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 10078–10086, 2023.

Wan, X., Liu, J., Gan, X., Liu, X., Wang, S., Wen, Y., Wan, T., and Zhu, E. One-step multi-view clustering with diverse representation. *IEEE Transactions on Neural Networks and Learning Systems*, 2024a.

Wan, X., Xiao, B., Liu, X., Liu, J., Liang, W., and Zhu, E. Fast continual multi-view clustering with incomplete views. *IEEE Transactions on Image Processing*, 2024b.

Wang, Q., Ding, Z., Tao, Z., Gao, Q., and Fu, Y. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30:1771–1783, 2021.

Wen, J., Xu, Y., and Liu, H. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Transactions on Cybernetics*, 50(4):1418–1429, 2020.

Wen, J., Sun, H., Fei, L., Li, J., Zhang, Z., and Zhang, B. Consensus guided incomplete multi-view spectral clustering. *Neural Networks*, 133:207–219, 2021a.

Wen, J., Zhang, Z., Zhang, Z., Zhu, L., Fei, L., Zhang, B., and Xu, Y. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 10273–10281, 2021b.

Wen, J., Liu, C., Xu, G., Wu, Z., Huang, C., Fei, L., and Xu, Y. Highly confident local structure based consensus graph learning for incomplete multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15712–15721, 2023.

Wong, W. K., Liu, C., Deng, S., Fei, L., Li, L., Lu, Y., and Wen, J. Neighbor group structure preserving based consensus graph learning for incomplete multi-view clustering. *Information Fusion*, 100:101917, 2023.

Xia, W., Gao, Q., Wang, Q., and Gao, X. Tensor completion-based incomplete multiview clustering. *IEEE Transactions on Cybernetics*, 52(12):13635–13644, 2022a.

Xia, W., Gao, Q., Wang, Q., Gao, X., Ding, C., and Tao, D. Tensorized bipartite graph learning for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022b.

Xu, J., Tang, H., Ren, Y., Peng, L., Zhu, X., and He, L. Multi-level feature learning for contrastive multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16051–16060, 2022.

Xu, J., Li, C., Peng, L., Ren, Y., Shi, X., Shen, H. T., and Zhu, X. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 32:1354–1366, 2023.

Xue, Z., Du, J., Zhu, H., Guan, Z., Long, Y., Zang, Y., and Liang, M. Robust diversified graph contrastive network for incomplete multi-view clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 3936–3944, 2022.

Yang, X., Deng, C., Dang, Z., and Tao, D. Deep multiview collaborative clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

Yu, S., Liu, S., Wang, S., Tang, C., Luo, Z., Liu, X., and Zhu, E. Sparse low-rank multi-view subspace clustering with consensus anchors and unified bipartite graph. *IEEE Transactions on Neural Networks and Learning Systems*, 2023a.

Yu, S., Wang, S., Wen, Y., Wang, Z., Luo, Z., Zhu, E., and Liu, X. How to construct corresponding anchors for incomplete multiview clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023b.

Yu, S., Wang, S., Dong, Z., Tu, W., Liu, S., Lv, Z., Li, P., Wang, M., and Zhu, E. A non-parametric graph clustering framework for multi-view data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16558–16567, 2024a.

Yu, S., Wang, S., Zhang, P., Wang, M., Wang, Z., Liu, Z., Fang, L., Zhu, E., and Liu, X. Dvsai: Diverse view-shared anchors based incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16568–16577, 2024b.

Zhang, C., Han, Z., Fu, H., Zhou, J. T., Hu, Q., et al. Cpm-nets: Cross partial multi-view networks. *Advances in Neural Information Processing Systems*, 32, 2019.

Zhang, Y., Liu, X., Wang, S., Liu, J., Dai, S., and Zhu, E. One-stage incomplete multi-view clustering via late fusion. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 2717–2725, 2021.

Zhao, H., Liu, H., and Fu, Y. Incomplete multi-modal visual data grouping. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 2392–2398, 2016.

## A. Proof of Theorem 3.1

*Proof.* Denote $\mathbf{h}_m = [\sum_{l=1}^{n_m}[\mathbf{W}_m]_{1,l}, \cdots, \sum_{l=1}^{n_m}[\mathbf{W}_m]_{k,l}, \cdots, \sum_{l=1}^{n_m}[\mathbf{W}_m]_{n,l}]_{1\times n}$. For the element $[\mathbf{A}_m]_{i,j}$, $i \neq j$, we have

$$\begin{aligned}
[\mathbf{A}_m]_{i,j} &= \mathbf{L}_{i,:}\mathbf{W}_m\mathbf{W}_m^\top[\mathbf{L}^\top]_{:,j} \\
&= \mathbf{L}_{i,:}diag(\mathbf{h}_m)[\mathbf{L}_{j,:}]^\top \\
&= \sum_{k=1}^{n}\mathbf{L}_{i,k}\mathbf{L}_{j,k}\sum_{l=1}^{n_m}[\mathbf{W}_m]_{k,l} \\
&= \sum_{k=1}^{n}0\sum_{l=1}^{n_m}[\mathbf{W}_m]_{k,l} \\
&= 0.
\end{aligned} \tag{22}$$

For $[\mathbf{A}_m]_{i,i}$, we have

$$\begin{aligned}
[\mathbf{A}_m]_{i,i} &= \sum_{k=1}^{n}\mathbf{L}_{i,k}\mathbf{L}_{i,k}\sum_{l=1}^{n_m}[\mathbf{W}_m]_{k,l} \\
&= \sum_{k=1}^{n}\mathbf{L}_{i,k}\sum_{l=1}^{n_m}[\mathbf{W}_m]_{k,l} \\
&= \mathbf{L}_{i,:}\mathbf{h}_m^\top.
\end{aligned} \tag{23}$$

Thus, we have that $\mathbf{A}_m$ is a diagonal matrix.

□

## B. Proof of Theorem 3.2

*Proof.* According to

$$f(\mathbf{O}) = \sum_{m=1}^{M}\mathrm{Tr}(\mathbf{O}^\top\widetilde{\mathbf{F}}\mathbf{O}\mathbf{A}_m + \mathbf{O}^\top\widetilde{\mathbf{F}}_m\mathbf{O}\mathbf{A}_m + \mathbf{O}^\top\mathbf{S}_m), \tag{24}$$

we have

$$\nabla f(\mathbf{O}) = \sum_{m=1}^{M}\widetilde{\mathbf{F}}\mathbf{O}\mathbf{A}_m + \widetilde{\mathbf{F}}^\top\mathbf{O}\mathbf{A}_m^\top + \widetilde{\mathbf{F}}_m\mathbf{O}\mathbf{A}_m + \widetilde{\mathbf{F}}_m^\top\mathbf{O}\mathbf{A}_m^\top + \mathbf{S}_m. \tag{25}$$

Given the fact that $\widetilde{\mathbf{F}} = \lambda_{max}(\mathbf{F})\mathbf{I}_c - \mathbf{F}$, $\mathbf{F} = \mathbf{E}^\top\mathbf{E}$, $\widetilde{\mathbf{F}}_m = \lambda_{max}(\mathbf{F}_m)\mathbf{I}_c - \mathbf{F}_m$, $\mathbf{F}_m = \mathbf{E}_m^\top\mathbf{E}_m$ and $\mathbf{A}_m = \mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top$, we know that $\widetilde{\mathbf{F}}$, $\widetilde{\mathbf{F}}_m$ and $\mathbf{A}_m$ are symmetric. Thus, we can further get

$$\nabla f(\mathbf{O}) = \sum_{m=1}^{M}2\widetilde{\mathbf{F}}\mathbf{O}\mathbf{A}_m + 2\widetilde{\mathbf{F}}_m\mathbf{O}\mathbf{A}_m + \mathbf{S}_m. \tag{26}$$

Denote the SVD of $\nabla f(\mathbf{O}_t)$ as $\mathbf{U}_t\mathbf{\Sigma}_t\mathbf{V}_t^\top$. We have

$$\mathrm{Tr}\left(\mathbf{O}^\top\nabla f(\mathbf{O}_t)\right) = \mathrm{Tr}\left(\mathbf{O}^\top\mathbf{U}_t\mathbf{\Sigma}_t\mathbf{V}_t^\top\right) = \mathrm{Tr}\left(\mathbf{V}_t^\top\mathbf{O}^\top\mathbf{U}_t\mathbf{\Sigma}_t\right) \leq \mathrm{Tr}\left(\mathbf{\Sigma}_t\right), \tag{27}$$

where the inequality holds based on the facts that $\mathbf{V}_t^\top\mathbf{O}^\top\mathbf{U}_t$ is an orthogonal matrix and that the diagonal elements of $\mathbf{\Sigma}_t$ is non-negative. Especially, when $\mathbf{V}_t^\top\mathbf{O}^\top\mathbf{U}_t = \mathbf{I}$, that is, $\mathbf{O} = \mathbf{U}_t\mathbf{V}_t^\top$, the third equality holds. Thus, for any $\mathbf{O}$, with $\mathbf{O}_{t+1} = \mathbf{U}_t\mathbf{V}_t$, we have

$$\mathrm{Tr}\left(\mathbf{O}^\top\nabla f(\mathbf{O}_t)\right) \leq \mathrm{Tr}\left(\mathbf{O}_{t+1}^\top\nabla f(\mathbf{O}_t)\right). \tag{28}$$

Specially, when $\mathbf{O}$ takes $\mathbf{O}_t$, Eq. (28) still holds. Thus, we have

$$\mathrm{Tr}\left(\mathbf{O}_t^\top\nabla f(\mathbf{O}_t)\right) \leq \mathrm{Tr}\left(\mathbf{O}_{t+1}^\top\nabla f(\mathbf{O}_t)\right). \tag{29}$$

□

## C. Proof of Theorem 3.3

*Proof.* Combined with Eqs. (26) and (29), we have

$$\mathrm{Tr}\left(\mathbf{O}_{t+1}^{\top}\left(\sum_{m=1}^{M}2\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m+2\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m+\mathbf{S}_m\right)\right)\geq\mathrm{Tr}\left(\mathbf{O}_{t}^{\top}\left(\sum_{m=1}^{M}2\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m+2\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m+\mathbf{S}_m\right)\right). \tag{30}$$

Combined with the definition of $f(\mathbf{O})$ in Eq. (24), we further have

$$\begin{aligned}① &\geq \sum_{m=1}^{M}\mathrm{Tr}\left(\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m+\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m+\mathbf{O}_t^{\top}\mathbf{S}_m\right)\\ &=f(\mathbf{O}_t),\end{aligned} \tag{31}$$

where

$$① = \sum_{m=1}^{M}\mathrm{Tr}\left(2\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m+2\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m-\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m-\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m+\mathbf{O}_{t+1}^{\top}\mathbf{S}_m\right). \tag{32}$$

Before proving $f(\mathbf{O}_{t+1})\geq f(\mathbf{O}_t)$, we only need to prove

$$f(\mathbf{O}_{t+1})\geq ①\,, \tag{33}$$

that is, prove that the following inequality holds

$$f(\mathbf{O}_{t+1})-①\geq 0. \tag{34}$$

Combined with Eqs. (24) and (32), we need to prove

$$② + ③ \geq 0, \tag{35}$$

where

$$② = \sum_{m=1}^{M}\mathrm{Tr}\left(\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}\mathbf{O}_{t+1}\mathbf{A}_m-2\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m+\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m\right)$$

and

$$③ = \sum_{m=1}^{M}\mathrm{Tr}\left(\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_{t+1}\mathbf{A}_m-2\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m+\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m\right).$$

.

Given that matrices $\mathbf{A}_m$ and $\widetilde{\mathbf{F}}$ are symmetric, we have

$$\mathrm{Tr}(\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m)=\mathrm{Tr}(\mathbf{A}_m^{\top}\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}^{\top}\mathbf{O}_{t+1})=\mathrm{Tr}(\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}^{\top}\mathbf{O}_{t+1}\mathbf{A}_m^{\top})=\mathrm{Tr}(\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}\mathbf{O}_{t+1}\mathbf{A}_m). \tag{36}$$

Thus, we have

$$\begin{aligned}② &= \sum_{m=1}^{M}\mathrm{Tr}\left(\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}\mathbf{O}_{t+1}\mathbf{A}_m-\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m-\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}\mathbf{O}_{t+1}\mathbf{A}_m+\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}\mathbf{O}_t\mathbf{A}_m\right)\\ &=\sum_{m=1}^{M}\mathrm{Tr}\left((\mathbf{O}_{t+1}-\mathbf{O}_t)^{\top}\widetilde{\mathbf{F}}(\mathbf{O}_{t+1}-\mathbf{O}_t)\mathbf{A}_m\right)\\ &\geq 0.\end{aligned} \tag{37}$$

Besides, $\widetilde{\mathbf{F}}_m$ is also symmetric, and therefore we can get

$$\mathrm{Tr}(\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m)=\mathrm{Tr}(\mathbf{O}_t^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_{t+1}\mathbf{A}_m). \tag{38}$$

Thus, for ③ , we have

$$\begin{aligned} ③ &= \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_{t+1}\mathbf{A}_m - \mathbf{O}_{t+1}^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m - \mathbf{O}_t^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_{t+1}\mathbf{A}_m + \mathbf{O}_t^{\top}\widetilde{\mathbf{F}}_m\mathbf{O}_t\mathbf{A}_m\right). \\ &= \sum_{m=1}^{M} \mathrm{Tr}\left((\mathbf{O}_{t+1}-\mathbf{O}_t)^{\top}\widetilde{\mathbf{F}}_m(\mathbf{O}_{t+1}-\mathbf{O}_t)\mathbf{A}_m\right) \\ &\geq 0. \end{aligned} \tag{39}$$

Combined with Eqs. (37) and (39), we have that Eq. (35) holds. That is, Eq. (34) holds and Eq. (33) holds.

Combined with Eqs. (33) and (31), we have

$$f(\mathbf{O}_{t+1}) \geq f(\mathbf{O}_t), \tag{40}$$

which indicates that under $\mathbf{O}_{t+1}$ taking $\mathbf{U}_t\mathbf{V}_t^{\top}$ where $\mathbf{U}_t$ and $\mathbf{V}_t^{\top}$ are the SVD results of $\nabla f(\mathbf{O}_t)$, $f(\mathbf{O})$ is non-decreasing.

□

## D. Proof of Theorem 3.4

*Proof.* For the optimization problem

$$\min_{\mathbf{G}_m\mathbf{G}_m^{\top}=\mathbf{I}_c} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2,$$

by expanding $F$-norm, we have

$$\begin{aligned} &\min_{\mathbf{G}_m} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 \Leftrightarrow \\ &\min_{\mathbf{G}_m} \mathrm{Tr}\left(\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^{\top}\mathbf{D}_m^{\top}\mathbf{G}_m^{\top} - 2\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^{\top}\mathbf{L}^{\top}\mathbf{O}^{\top}\mathbf{E}^{\top} + \mathbf{EOLW}_m\mathbf{W}_m^{\top}\mathbf{L}^{\top}\mathbf{O}^{\top}\mathbf{E}^{\top}\right) \Leftrightarrow \\ &\min_{\mathbf{G}_m} \mathrm{Tr}\left(\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^{\top}\mathbf{D}_m^{\top}\mathbf{G}_m^{\top} - 2\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^{\top}\mathbf{L}^{\top}\mathbf{O}^{\top}\mathbf{E}^{\top}\right) \Leftrightarrow \\ &\min_{\mathbf{G}_m} \mathrm{Tr}\left(\mathbf{G}_m\mathbf{P}_m\mathbf{G}_m^{\top} - \mathbf{G}_m\mathbf{Z}_m^{\top}\right), \end{aligned} \tag{41}$$

where $\mathbf{P}_m = \mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^{\top}\mathbf{D}_m^{\top}$, $\mathbf{Z}_m = 2\mathbf{EOLW}_m\mathbf{W}_m^{\top}\mathbf{D}_m^{\top}$.

Combined with $\mathbf{G}_m\mathbf{G}_m^{\top} = \mathbf{I}_c$, we further have

$$\begin{aligned} &\min_{\mathbf{G}_m} \mathrm{Tr}\left(\mathbf{G}_m\mathbf{P}_m\mathbf{G}_m^{\top} - \mathbf{G}_m\mathbf{Z}_m^{\top}\right) \Leftrightarrow \\ &\min_{\mathbf{G}_m} \mathrm{Tr}\left(\mathbf{G}_m\left(\mathbf{P}_m - \lambda_{max}(\mathbf{P}_m)\mathbf{I}_{d_m}\right)\mathbf{G}_m^{\top} - \mathbf{G}_m\mathbf{Z}_m^{\top}\right) \Leftrightarrow \\ &\max_{\mathbf{G}_m} \mathrm{Tr}\left(\mathbf{G}_m\widetilde{\mathbf{P}}_m\mathbf{G}_m^{\top} + \mathbf{G}_m\mathbf{Z}_m^{\top}\right). \end{aligned} \tag{42}$$

where $\widetilde{\mathbf{P}}_m = \lambda_{max}(\mathbf{P}_m)\mathbf{I}_{d_m} - \mathbf{P}_m$.

Denote $g(\mathbf{G}_m) = \mathrm{Tr}\left(\mathbf{G}_m\widetilde{\mathbf{P}}_m\mathbf{G}_m^{\top} + \mathbf{G}_m\mathbf{Z}_m^{\top}\right)$. Its derivative $\nabla g\left(\mathbf{G}_m\right)$ is $\mathbf{G}_m\widetilde{\mathbf{P}}_m^{\top} + \mathbf{G}_m\widetilde{\mathbf{P}}_m + \mathbf{Z}_m$. Thus, the solution of $\mathbf{G}_m$ can be set as $\mathbf{UV}^{\top}$, where $\mathbf{U}$ and $\mathbf{V}$ are the SVD results of $[\mathbf{G}_m]_{pre}\widetilde{\mathbf{P}}_m^{\top} + [\mathbf{G}_m]_{pre}\widetilde{\mathbf{P}}_m + \mathbf{Z}_m$. $[\mathbf{G}_m]_{pre}$ represents the value of $\mathbf{G}_m$ at previous iteration. Considering that $\widetilde{\mathbf{P}}_m$ is a symmetric matrix, thus, we can perform SVD on $2[\mathbf{G}_m]_{pre}\widetilde{\mathbf{P}}_m + \mathbf{Z}_m$ to generate $\mathbf{U}$ and $\mathbf{V}$, and thereby obtain the solution of $\mathbf{G}_m$.

□

## E. Proof of Theorem 3.5

*Proof.* The computational cost of the proposed algorithm ToRES is mainly composed of optimizing $\mathbf{O}$, $\mathbf{L}$, $\mathbf{G}_m$, $\mathbf{E}$ and $\mathbf{E}_m$.

When updating $\mathbf{O}$ using Algorithm 1, the construction of $\mathbf{A}_m = \mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top = (\mathbf{L} \odot \mathbf{Q}_m)\,\mathbf{L}^\top$ will take $\mathcal{O}(cn + c^2n)$. $\mathbf{B}_m = \mathbf{E}^\top\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top = \mathbf{E}^\top\mathbf{G}_m\,(\mathbf{D}_m \odot \mathbf{H}_m)\,\mathbf{L}^\top$ will take $\mathcal{O}(d_mn + c^2d_m + cd_mn + c^2n)$. $\mathbf{C}_m = \mathbf{E}_m^\top\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top = \mathbf{E}_m^\top\,(\mathbf{D}_m \odot \mathbf{H}_m)\,\mathbf{L}^\top$ will take $\mathcal{O}(d_mn + cd_mn + c^2n)$. $\mathbf{Q}_m \in \mathbb{R}^{c\times n} = \mathbf{1}_{c\times 1}\cdot\mathbf{h}_m$, $\mathbf{H}_m \in \mathbb{R}^{d_m\times n} = \mathbf{1}_{d_m\times 1}\cdot\mathbf{h}_m$, $\mathbf{h}_m = [\sum_{l=1}^{n_m}[\mathbf{W}_m]_{1,l}, \cdots, \sum_{l=1}^{n_m}[\mathbf{W}_m]_{k,l}, \cdots, \sum_{l=1}^{n_m}[\mathbf{W}_m]_{n,l}]_{1\times n}$. The constructions of $\mathbf{E}^\top\mathbf{E}$ and $\mathbf{E}_m^\top\mathbf{E}_m$ will take $\mathcal{O}(c^3)$ and $\mathcal{O}(c^2d_m)$ respectively. Using quadratic programming with quadratic constraints to solve $\mathbf{O}$ will take $\mathcal{O}(c^4)$. Thus, the computational complexity of updating $\mathbf{O}$ by Algorithm 1 is $\mathcal{O}(c^2d + cdn + c^2Mn + c^4)$, where $d = \sum_{m=1}^{M} d_m$.

When updating $\mathbf{O}$ using Algorithm 2, the constructions of $\widetilde{\mathbf{F}}$, $\widetilde{\mathbf{F}}_m$, $\mathbf{A}_m$ and $\mathbf{S}_m$ will take $\mathcal{O}(c^3)$, $\mathcal{O}(c^2d_m + c^3)$, $\mathcal{O}(cn + c^2n)$ and $\mathcal{O}(d_mn + c^2d_m + cd_mn + c^2n)$ respectively. Calculating $\nabla f(\mathbf{O})$ will take $\mathcal{O}(c^3)$. The singular value decomposition and singular matrix multiplication of $\nabla f(\mathbf{O})$ will take both $\mathcal{O}(c^3)$. Thus, the computational complexity of updating $\mathbf{O}$ by Algorithm 2 is $\mathcal{O}(cdn + c^2nM + c^2d + Mc^3)$.

When updating $\mathbf{L}$, the constructions of $\mathbf{T}_j$ and $\mathbf{J}$ will take $\mathcal{O}(c^3 + c^2d)$ and $\mathcal{O}(ndc + nc^2M)$ respectively. Thus, the computational complexity of updating $\mathbf{L}$ is $\mathcal{O}(c^3n + c^2dn + nc^2M)$.

When updating $\mathbf{G}_m$, constructing $\widetilde{\mathbf{P}}_m$, $(2[\mathbf{G}_m]_{pre}\widetilde{\mathbf{P}}_m + \mathbf{Z}_m)$ and performing SVD on it will take $\mathcal{O}(d_mn + d_m^2n + d_m^3)$, $\mathcal{O}(cd_m^2)$ and $\mathcal{O}(c^2d_m)$ respectively. Thus, updating $\mathbf{G}_m$ will take $\mathcal{O}(d_m^2n + d_m^3 + c^2d_m)$. The computational complexity of updating all $\mathbf{G}_m$ is $\mathcal{O}(n\sum_{m=1}^{M} d_m^2 + \sum_{m=1}^{M} d_m^3 + c^2d)$.

When updating $\mathbf{E}$, constructing $\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top$, $\mathbf{O}\sum_{m=1}^{M}\mathbf{A}_m\mathbf{O}^\top$ and its inverse will take $\mathcal{O}(d_mn + cd_mn)$, $\mathcal{O}(c^3)$ and $\mathcal{O}(c^3)$ respectively. Thus, the computational complexity of updating $\mathbf{E}$ is $\mathcal{O}(cdn + c^2n + c^3)$.

When updating $\mathbf{E}_m$, it will take $\mathcal{O}(d_mnc + d_mc^2 + c^3)$. Thus, the computational complexity of updating all $\mathbf{E}_m$ is $\mathcal{O}(dnc + dc^2 + Mc^3)$.

In addition, commonly, $c \ll n$, $M \ll n$, and the data dimension $d_m$ is a constant and independent of the sample number $n$.

Therefore, we have that the computational complexity of ToRES is $\mathcal{O}(c^3n + c^2Mn + c^2dn + n\sum_{m=1}^{M} d_m^2 + \sum_{m=1}^{M} d_m^3 + c^4)$ when using Algorithm 1. When using Algorithm 2, the computational complexity of ToRES is $\mathcal{O}(c^2nM + Mc^3 + c^3n + c^2dn + n\sum_{m=1}^{M} d_m^2 + \sum_{m=1}^{M} d_m^3)$. They are both linear with respect to the number of samples $n$. Thus, the computational complexity of our ToRES is $\mathcal{O}(n)$. □

## F. Detailed Derivations for Optimization Variables $\mathbf{G}_m$, $\mathbf{E}$, $\mathbf{O}$, $\mathbf{L}$ and $\mathbf{E}_m$

In this section, we give detailed derivations about $\mathbf{G}_m$, $\mathbf{E}$, $\mathbf{O}$, $\mathbf{L}$ and $\mathbf{E}_m$.

★ ***Step-One:*** *Optimizing* $\mathbf{O}$.

With other variables fixed, Eq. (3) about $\mathbf{O}$ becomes

$$\min_{\mathbf{O}^\top\mathbf{O}=\mathbf{I}_c} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 + \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 .$$

**Scheme 1:**

For $\|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2$, we have

$$\begin{aligned}
&\min_{\mathbf{O}} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{O}} \mathrm{Tr}\left(\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{G}_m^\top - 2\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top + \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top\right) \Leftrightarrow \\
&\min_{\mathbf{O}} \mathrm{Tr}\left(\mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top - 2\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top\right) \Leftrightarrow \\
&\min_{\mathbf{O}} \mathrm{Tr}\left(\mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top - 2\mathbf{E}^\top\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\right) \Leftrightarrow \\
&\min_{\mathbf{O}} \mathrm{Tr}\left(\mathbf{O}^\top\mathbf{E}^\top\mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top - 2\left[\mathbf{E}^\top\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\right]^\top\mathbf{O}\right) \Leftrightarrow \\
&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}^\top\mathbf{E}\mathbf{O}[\mathbf{A}_m]_{:,j} - 2[\mathbf{B}_m]_{:,j}^\top\mathbf{O}_{:,j},
\end{aligned} \tag{43}$$

where $\mathbf{A}_m = \mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top$, $\mathbf{B}_m = \mathbf{E}^\top\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top$.

According to **Theorem 3.1**, we can get

$$\begin{aligned}
&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}^\top\mathbf{E}\mathbf{O}[\mathbf{A}_m]_{:,j} - 2[\mathbf{B}_m]_{:,j}^\top\mathbf{O}_{:,j} \Leftrightarrow \\
&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}^\top\mathbf{E}[\mathbf{A}_m]_{j,j}\mathbf{O}_{:,j} - 2[\mathbf{B}_m]_{:,j}^\top\mathbf{O}_{:,j}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\min_{\mathbf{O}} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}^\top\mathbf{E}[\mathbf{A}_m]_{j,j}\mathbf{O}_{:,j} - 2[\mathbf{B}_m]_{:,j}^\top\mathbf{O}_{:,j}.
\end{aligned} \tag{44}$$

For $\|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2$, we have

$$\begin{aligned}
&\min_{\mathbf{O}} \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{O}} \mathrm{Tr}\left(\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top - 2\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top + \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top\right) \Leftrightarrow \\
&\min_{\mathbf{O}} \mathrm{Tr}\left(\mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top - 2\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top\right) \Leftrightarrow \\
&\min_{\mathbf{O}} \mathrm{Tr}\left(\mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top - 2\mathbf{E}_m^\top\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\right) \Leftrightarrow \\
&\min_{\mathbf{O}} \mathrm{Tr}\left(\mathbf{O}^\top\mathbf{E}_m^\top\mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top - 2\left[\mathbf{E}_m^\top\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\right]^\top\mathbf{O}\right) \Leftrightarrow \\
&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}_m^\top\mathbf{E}_m\mathbf{O}[\mathbf{A}_m]_{:,j} - 2[\mathbf{C}_m]_{:,j}^\top\mathbf{O}_{:,j} \Leftrightarrow \\
&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\mathbf{E}_m^\top\mathbf{E}_m[\mathbf{A}_m]_{j,j}\mathbf{O}_{:,j} - 2[\mathbf{C}_m]_{:,j}^\top\mathbf{O}_{:,j},
\end{aligned} \tag{45}$$

where $\mathbf{C}_m = \mathbf{E}_m^\top\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top$.

Thus, we can get

$$\begin{aligned}
&\min_{\mathbf{O}} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 + \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{O}_{:,j}} \sum_{m=1}^{M} \mathbf{O}_{:,j}^\top\mathbf{E}^\top\mathbf{E}[\mathbf{A}_m]_{j,j}\mathbf{O}_{:,j} - 2[\mathbf{B}_m]_{:,j}^\top\mathbf{O}_{:,j} + \mathbf{O}_{:,j}^\top\mathbf{E}_m^\top\mathbf{E}_m[\mathbf{A}_m]_{j,j}\mathbf{O}_{:,j} - 2[\mathbf{C}_m]_{:,j}^\top\mathbf{O}_{:,j} \Leftrightarrow \\
&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\left(\mathbf{E}^\top\mathbf{E}\sum_{m=1}^{M}[\mathbf{A}_m]_{j,j} + \sum_{m=1}^{M}\mathbf{E}_m^\top\mathbf{E}_m[\mathbf{A}_m]_{j,j}\right)\mathbf{O}_{:,j} + 2\left(-\sum_{m=1}^{M}\mathbf{B}_m - \sum_{m=1}^{M}\mathbf{C}_m\right)_{:,j}^\top\mathbf{O}_{:,j}.
\end{aligned} \tag{46}$$

Considering that the feasible region $\mathbf{O}^\top\mathbf{O} = \mathbf{I}_c$ is equivalent to $\mathbf{O}_{:,j}^\top\mathbf{O}_{:,j} = 1$, $\mathbf{O}_{:,j}^\top\mathbf{O}_{:,i} = 0$, $i = [1, 2, \cdots, c]$ and $i \neq j$. $j = 1, 2, \cdots c$. Therefore, we have

$$\begin{aligned}
&\min_{\mathbf{O}^\top\mathbf{O}=\mathbf{I}_c} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 + \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{O}_{:,j}} \mathbf{O}_{:,j}^\top\left(\mathbf{E}^\top\mathbf{E}\sum_{m=1}^{M}[\mathbf{A}_m]_{j,j} + \sum_{m=1}^{M}\mathbf{E}_m^\top\mathbf{E}_m[\mathbf{A}_m]_{j,j}\right)\mathbf{O}_{:,j} + 2\left(-\sum_{m=1}^{M}\mathbf{B}_m - \sum_{m=1}^{M}\mathbf{C}_m\right)_{:,j}^\top\mathbf{O}_{:,j} \\
&\quad \text{s.t. } \frac{1}{2}\mathbf{O}_{:,j}^\top\mathbf{I}\mathbf{O}_{:,j} + \left(-\frac{1}{2}\right) = 0,\ \left[\mathbf{O}_{:,1}, \mathbf{O}_{:,2}, \cdots, \mathbf{O}_{:,j-1}, \mathbf{O}_{:,j+1} \cdots \mathbf{O}_{:,c}\right]^\top\mathbf{O}_{:,j} = \mathbf{0}.
\end{aligned} \tag{47}$$

**Scheme 2:**

In conjunction with Eqs. (43) and (45), we can get

$$\begin{aligned}
&\min_{\mathbf{O}} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 + \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{O}} \sum_{m=1}^{M} \mathrm{Tr}\Big(\mathbf{O}^\top\mathbf{E}^\top\mathbf{EOLW}_m\mathbf{W}_m^\top\mathbf{L}^\top - 2\left[\mathbf{E}^\top\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\right]^\top\mathbf{O} + \\
&\qquad\qquad \mathbf{O}^\top\mathbf{E}_m^\top\mathbf{E}_m\mathbf{OLW}_m\mathbf{W}_m^\top\mathbf{L}^\top - 2\left[\mathbf{E}_m^\top\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\right]^\top\mathbf{O}\Big) \Leftrightarrow \\
&\min_{\mathbf{O}} \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{O}^\top\mathbf{E}^\top\mathbf{EOA}_m + \mathbf{O}^\top\mathbf{E}_m^\top\mathbf{E}_m\mathbf{OA}_m - 2\mathbf{B}_m^\top\mathbf{O} - 2\mathbf{C}_m^\top\mathbf{O}\right) \Leftrightarrow \\
&\min_{\mathbf{O}} \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{O}^\top\mathbf{FOA}_m + \mathbf{O}^\top\mathbf{F}_m\mathbf{OA}_m - \mathbf{O}^\top\mathbf{S}_m\right),
\end{aligned} \tag{48}$$

where $\mathbf{F} = \mathbf{E}^\top\mathbf{E}$, $\mathbf{F}_m = \mathbf{E}_m^\top\mathbf{E}_m$, $\mathbf{S}_m = 2\left(\mathbf{B}_m + \mathbf{C}_m\right)$, $\mathbf{B}_m = \mathbf{E}^\top\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top$, $\mathbf{C}_m = \mathbf{E}_m^\top\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top$.

Due to $\mathbf{O}^\top\mathbf{O} = \mathbf{I}_c$, we further have

$$\begin{aligned}
&\min_{\mathbf{O}} \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{O}^\top\mathbf{FOA}_m + \mathbf{O}^\top\mathbf{F}_m\mathbf{OA}_m - \mathbf{O}^\top\mathbf{S}_m\right) \Leftrightarrow \\
&\min_{\mathbf{O}} \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{O}^\top\left(\mathbf{F} - \lambda_{max}(\mathbf{F})\mathbf{I}_c\right)\mathbf{OA}_m + \mathbf{O}^\top\left(\mathbf{F}_m - \lambda_{max}(\mathbf{F}_m)\mathbf{I}_c\right)\mathbf{OA}_m - \mathbf{O}^\top\mathbf{S}_m\right) \Leftrightarrow \\
&\max_{\mathbf{O}} \sum_{m=1}^{M} \mathrm{Tr}\left(\mathbf{O}^\top\widetilde{\mathbf{F}}\mathbf{OA}_m + \mathbf{O}^\top\widetilde{\mathbf{F}}_m\mathbf{OA}_m + \mathbf{O}^\top\mathbf{S}_m\right),
\end{aligned} \tag{49}$$

where $\widetilde{\mathbf{F}} = \lambda_{max}(\mathbf{F})\mathbf{I}_c - \mathbf{F}$, $\widetilde{\mathbf{F}}_m = \lambda_{max}(\mathbf{F}_m)\mathbf{I}_c - \mathbf{F}_m$.

Combined with $f(\mathbf{O}) = \sum_{m=1}^{M}\mathrm{Tr}(\mathbf{O}^\top\widetilde{\mathbf{F}}\mathbf{OA}_m + \mathbf{O}^\top\widetilde{\mathbf{F}}_m\mathbf{OA}_m + \mathbf{O}^\top\mathbf{S}_m)$ and Theorem 3.3, we know that under $\mathbf{O}_{t+1}$ taking $\mathbf{U}_t\mathbf{V}_t^\top$ where $\mathbf{U}_t$ and $\mathbf{V}_t^\top$ are SVD results of $\nabla f(\mathbf{O}_t)$, the function $f(\mathbf{O})$ is non-decreasing. Meanwhile, the solution $\mathbf{O}_{t+1} = \mathbf{U}_t\mathbf{V}_t^\top$ lies in the flexible region $\mathbf{O}^\top\mathbf{O} = \mathbf{I}_c$. Thus, we can take $\mathbf{O}_{t+1}$ as the solution of $\mathbf{O}$ when $\mathbf{O}_{t+1}$ and $\mathbf{O}_t$ satisfy the condition $\|\mathbf{O}_{t+1} - \mathbf{O}_t\|_F / \|\mathbf{O}_t\|_F \leq 1e-5$.

★ ***Step-Two:*** *Optimizing* $\mathbf{L}$.

With other variables fixed, Eq. (3) about $\mathbf{L}$ becomes

$$\begin{aligned}
&\min_{\mathbf{L}} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 + \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2 \\
&\qquad \text{s.t. } \mathbf{L} \in \{0,1\}^{c\times n}, \|\mathbf{L}_{:,j}\|_1 = 1, j \in \{1,2,\ldots,n\}.
\end{aligned}$$

Combined with Eq. (43), we have

$$\begin{aligned}
&\min_{\mathbf{L}} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{L}} \mathrm{Tr}\left(\mathbf{EOLW}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top - 2\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top\right) \Leftrightarrow \\
&\min_{\mathbf{L}} \mathrm{Tr}\left(\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top\mathbf{EOLW}_m\mathbf{W}_m^\top - 2\mathbf{O}^\top\mathbf{E}^\top\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\right) \Leftrightarrow \\
&\min_{\mathbf{L}} \mathrm{Tr}\left(\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top\mathbf{EOLW}_m\mathbf{W}_m^\top - 2\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{G}_m^\top\mathbf{EOL}\right).
\end{aligned} \tag{50}$$

Considering that

$$\mathbf{W}_m\mathbf{W}_m^\top = diag\left(\left[\sum_{l=1}^{n_m}[\mathbf{W}_m]_{1,l}, \cdots, \sum_{l=1}^{n_m}[\mathbf{W}_m]_{k,l}, \cdots, \sum_{l=1}^{n_m}[\mathbf{W}_m]_{n,l}\right]_{1\times n}\right), \tag{51}$$

we can have that

$$\begin{aligned}&\min_{\mathbf{L}} \operatorname{Tr}\left(\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}^\top\mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top - 2\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{G}_m^\top\mathbf{E}\mathbf{O}\mathbf{L}\right) \Leftrightarrow \\ &\min_{\mathbf{L}_{:,j}} \mathbf{L}_{:,j}^\top\mathbf{O}^\top\mathbf{E}^\top\mathbf{E}\mathbf{O}\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k}\mathbf{L}_{:,j} - 2\left(\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{G}_m^\top\mathbf{E}\mathbf{O}\right)_{j,:}\mathbf{L}_{:,j},\end{aligned} \tag{52}$$

Combined with Eqs. (50) and (52), we have

$$\begin{aligned}&\min_{\mathbf{L}} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\ &\min_{\mathbf{L}_{:,j}} \mathbf{L}_{:,j}^\top\mathbf{O}^\top\mathbf{E}^\top\mathbf{E}\mathbf{O}\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k}\mathbf{L}_{:,j} - 2\left(\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{G}_m^\top\mathbf{E}\mathbf{O}\right)_{j,:}\mathbf{L}_{:,j}.\end{aligned} \tag{53}$$

Additionally, based on Eq. (45), we have

$$\begin{aligned}&\min_{\mathbf{L}} \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\ &\min_{\mathbf{L}} \operatorname{Tr}\left(\mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top - 2\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top\right) \Leftrightarrow \\ &\min_{\mathbf{L}} \operatorname{Tr}\left(\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top\mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top - 2\mathbf{O}^\top\mathbf{E}_m^\top\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\right) \Leftrightarrow \\ &\min_{\mathbf{L}} \operatorname{Tr}\left(\mathbf{L}^\top\mathbf{O}^\top\mathbf{E}_m^\top\mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top - 2\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{E}_m\mathbf{O}\mathbf{L}\right) \Leftrightarrow \\ &\min_{\mathbf{L}_{:,j}} \mathbf{L}_{:,j}^\top\mathbf{O}^\top\mathbf{E}_m^\top\mathbf{E}_m\mathbf{O}\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k}\mathbf{L}_{:,j} - 2\left(\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{E}_m\mathbf{O}\right)_{j,:}\mathbf{L}_{:,j}.\end{aligned} \tag{54}$$

Combined with Eqs. (53) and (54), we have

$$\begin{aligned}&\min_{\mathbf{L}} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 + \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\ &\min_{\mathbf{L}_{:,j}} \sum_{m=1}^{M} \mathbf{L}_{:,j}^\top\mathbf{O}^\top\mathbf{E}^\top\mathbf{E}\mathbf{O}\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k}\mathbf{L}_{:,j} - 2\left(\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{G}_m^\top\mathbf{E}\mathbf{O}\right)_{j,:}\mathbf{L}_{:,j} + \\ &\qquad \mathbf{L}_{:,j}^\top\mathbf{O}^\top\mathbf{E}_m^\top\mathbf{E}_m\mathbf{O}\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k}\mathbf{L}_{:,j} - 2\left(\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\mathbf{E}_m\mathbf{O}\right)_{j,:}\mathbf{L}_{:,j} \Leftrightarrow \\ &\min_{\mathbf{L}_{:,j}} \mathbf{L}_{:,j}^\top\left(\mathbf{O}^\top\sum_{m=1}^{M}\left(\mathbf{E}^\top\mathbf{E}\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k} + \mathbf{E}_m^\top\mathbf{E}_m\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k}\right)\mathbf{O}\right)\mathbf{L}_{:,j} \\ &\qquad - 2\left(\sum_{m=1}^{M}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\left(\mathbf{G}_m^\top\mathbf{E} + \mathbf{E}_m\right)\mathbf{O}\right)_{j,:}\mathbf{L}_{:,j}.\end{aligned} \tag{55}$$

Thus, we have

$$\begin{aligned}&\min_{\mathbf{L}\in\{0,1\}^{c\times n},\|\mathbf{L}_{:,j}\|_1=1,j\in\{1,2,\ldots,n\}} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{E}\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 + \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\|_F^2 \Leftrightarrow \\ &\min_{\mathbf{L}_{:,j}\in\{0,1\}^{c\times 1},\|\mathbf{L}_{:,j}\|_1=1,j\in\{1,2,\ldots,n\}} \mathbf{L}_{:,j}^\top\mathbf{T}_j\mathbf{L}_{:,j} - \mathbf{J}_{j,:}\mathbf{L}_{:,j},\end{aligned} \tag{56}$$

where

$$\mathbf{T}_j = \mathbf{O}^\top\sum_{m=1}^{M}\left(\mathbf{E}^\top\mathbf{E}\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k} + \mathbf{E}_m^\top\mathbf{E}_m\sum_{k=1}^{n_m}[\mathbf{W}_m]_{j,k}\right)\mathbf{O}$$

$$\mathbf{J} = 2\sum_{m=1}^{M}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{D}_m^\top\left(\mathbf{G}_m^\top\mathbf{E} + \mathbf{E}_m\right)\mathbf{O}.$$

The constraints $\mathbf{L} \in \{0,1\}^{c\times n}, \|\mathbf{L}_{:,j}\|_1 = 1, j \in \{1,2,\ldots,n\}$ indicate that there is only one non-zero element in each column. Meanwhile, the operation $\mathbf{L}_{:,j}^\top \mathbf{T}_j \mathbf{L}_{:,j}$ means taking the diagonal elements of $\mathbf{T}_j \in \mathbb{R}^{c\times c}$. Therefore, we can determine $\mathbf{L}_{:,j}$ by comparing the diagonal elements of $\mathbf{T}_j$ and the $j$-th row of $\mathbf{J} \in \mathbb{R}^{n\times c}$. When $i = i^*$, $\mathbf{L}_{i,j}$ is set as 1, otherwise 0, where

$$i^* = \arg\min_i \ [\mathbf{T}_j]_{i,i} - \mathbf{J}_{j,i}, i \in \{1,2,\cdots,c\}.$$

★ ***Step-Three:*** *Optimizing* $\mathbf{G}_m$.

With other variables fixed, Eq. (3) about $\mathbf{G}_m$ becomes

$$\min_{\mathbf{G}_m\mathbf{G}_m^\top=\mathbf{I}_c} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 . \tag{57}$$

Due to $\{\mathbf{G}_m\}_{m=1}^M$ aiming at projecting respective views and being related to views, thus $\{\mathbf{G}_m\}_{m=1}^M$ are mutually independent. On basis of this, we can equivalently transform Eq. (57) as

$$\min_{\mathbf{G}_m\mathbf{G}_m^\top=\mathbf{I}_c} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 .$$

In virtue of Theorem 3.4, we can obtain the solution of $\mathbf{G}_m$.

★ ***Step-Four:*** *Optimizing* $\mathbf{E}$.

With other variables fixed, Eq. (3) about $\mathbf{E}$ becomes

$$\min_{\mathbf{E}} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\|_F^2 .$$

Since this is an unconstrained optimization problem, we can get the optimal solution by setting its derivative to zero. Thus, we have

$$\sum_{m=1}^{M} \left(\mathbf{G}_m\mathbf{D}_m\mathbf{W}_m - \mathbf{EOLW}_m\right)\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top = \mathbf{0}.$$

That is, we have

$$\mathbf{EOL}\sum_{m=1}^{M} \mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top = \sum_{m=1}^{M} \mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top. \tag{58}$$

Due to $\mathbf{A}_m = \mathbf{LW}_m\mathbf{W}_m^\top\mathbf{L}^\top$ being reversible, we can get

$$\mathbf{E} = \left(\sum_{m=1}^{M} \mathbf{G}_m\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\right)\mathbf{L}^\top\mathbf{O}^\top\left(\mathbf{O}\sum_{m=1}^{M}\mathbf{A}_m\mathbf{O}^\top\right)^{-1}.$$

★ ***Step-Five:*** *Optimizing* $\mathbf{E}_m$.

With other variables fixed, Eq. (3) about $\mathbf{E}_m$ becomes

$$\min_{\mathbf{E}_m} \sum_{m=1}^{M} \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2 . \tag{59}$$

View-wise prototypes $\{\mathbf{E}_m\}_{m=1}^M$ are related to individual views, and thus are independent of each other. Eq. (59) can be simplified as

$$\min_{\mathbf{E}_m} \|\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\|_F^2 .$$

It is an unconstrained optimization problem, and therefore, by setting its derivative to zero, we can get

$$\left(\mathbf{D}_m\mathbf{W}_m - \mathbf{E}_m\mathbf{OLW}_m\right)\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top = \mathbf{0}. \tag{60}$$

That is,

$$\mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top = \mathbf{E}_m\mathbf{O}\mathbf{L}\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top. \tag{61}$$

Thus, the optimal solution is

$$\mathbf{E}_m = \mathbf{D}_m\mathbf{W}_m\mathbf{W}_m^\top\mathbf{L}^\top\mathbf{O}^\top\left(\mathbf{O}\mathbf{A}_m\mathbf{O}^\top\right)^{-1}.$$

## G. More Related Work

For effectively grouping multi-view data, a series of methods have been proposed recently. For instances, Ma et al. (2024) concurrently perform the symmetrization and localization of affinity matrix so as to discover good neighbors automatically, and exploits local structure of unified affinity without needing conventional rank constraints and extra pre-searching hyper-parameters. Wan et al. (2024a) generate diverse representation by mapping original samples into multiple potential embedding spaces, and unify representation learning and $k$-means together so as to get rid of the impact of the procedure separation. Liang et al. (2022a) establish the generalization bound by approximating the eigen functions of basic kernel matrices to efficiently tackle out-of-sample points, and provide a uniform stability for the single-view scenarios. Li et al. (2019) introduce hierarchical self-expressive layers to recover potential low-dimensional spaces hierarchically, and exploit complicated relationships between views via a reciprocal subspace learning to encode multi-view information. Wan et al. (2022) provide a continual clustering scheme to deal with new view data, and aggregate all view information via a late fusion strategy. Liang et al. (2024) maintain the consistency of kernel parameters by building up the excess clustering risk, and decrease the computing overhead by substituting for the eigen decomposition using SVD. Li et al. (2023b) describe the cluster distributions by approximating the rank of 3rd-order Schatten p-norm to exploit knowledge between views, and utilize one-side constraint to consider the spatial structure embedded into each view. Wan et al. (2023) automatically weight base matrices with various dimensions to address the dimension-fixed issue, and map original data into various low-dimensional spaces to produce comprehensive representation. Liang et al. (2022b) introduce a min-max paradigm to perform robust learning, and utilize the reduced gradient descent strategy to optimize the convex objective. Pan & Kang (2021) employ graph filtering to clear high-frequency noise so as to maintain the geometric features, and construct the consensus graph via contrastive learning to decrease the impact of incomplete graphs. Other methods like (Li et al., 2024; Fu et al., 2023; Li et al., 2023c; Liang et al., 2022c; Liu et al., 2022b) are also well investigated.

## H. The URLs of Benchmark Datasets

In this subsection, we provide the URLs of Benchmark Datasets used in Table 3.

**Small-scale Datasets:**

Webkb: `https://www.cs.cmu.edu/~webkb/`

Wikifea: `http://svcl.ucsd.edu/projects/crossmodal/`

**Middle-scale Datasets:**

AWA10: `https://cvml.ista.ac.at/AwA2/`

SUNRGB-D: `https://rgbd.cs.princeton.edu/`

**Large-scale Datasets:**

AwAfea: `https://cvml.ista.ac.at/AwA/`

EMNIST: `https://www.nist.gov/itl/products-and-services/emnist-dataset`

Their brief description is as follows:

1. **Webkb**: It is sampled from diverse web-pages, and consists of 1051 samples. The cluster number is 2, and the data dimensions on 2 views are 334 and 2949.

2. **Wikifea**: It is extracted from multimedia documents, and contains 2866 samples. The cluster number is 10, and the data dimensions on 2 views are 128 and 10.

3. **AWA10**: This image dataset characterizes animals with various attributes, and has 5814 samples. There are 10 clusters. The data dimensions on 6 views are 2688, 2000, 252, 2000, 2000 and 2000 respectively.
4. **SUNRGB-D**: This image dataset is collected through RGB-D sensors in scene understanding, and has 10335 samples. There are 45 clusters. The data dimensions on 2 views are both 4096.
5. **AwAfea**: It describes different animal properties via a group of image feature extractors. The numbers of samples and clusters are 30475 and 50 respectively. The data dimensions on 6 views are 2688, 2000, 252, 2000, 2000 and 2000.
6. **EMNIST**: It is composed of different handwritten character digits. The numbers of samples and clusters are 280000 and 10. The feature dimensions on 4 views are 944, 512, 576 and 640 respectively.

## I. Detailed Descriptions for 20 Comparison Algorithms

The detailed descriptions of these 20 comparison algorithms are as follows:

1. **IMSC-AGL (Wen et al., 2020).** This algorithm integrates spectral learning and graph learning techniques to capture view-shared features for incomplete data, and mines consensus low-rank characteristics by introducing a co-regularization item to discover the intrinsic space structure among samples.
2. **AWP (Nie et al., 2018).** This algorithm introduces adaptive Procrustes Average scheme into spectral rotation to relieve the negative factor of clustering capacity differences between views, and decreases the disagreements by assessing views using a consistent indicator rather than individually to increase its applicability.
3. **APMC (Guo & Ye, 2019).** This algorithm combines inter- and intra- view affinity information via a group of landmarks to sufficiently extract non-linear relations between instances, and generates the class indicator by performing spectral clustering on the merged similarity graph under the help of Gaussian kernel function.
4. **IMG (Zhao et al., 2016).** This algorithm transforms multi-modal samples to a complete representation space instead of projecting views into the common subspace, and builds up the connection between partial-view instances and complete-view instances via a Laplacian graph to preserve tight global structure.
5. **TMBSD (Li et al., 2021c).** This algorithm preserves global block-diagonal structure across views using several spectral embedding matrices under the guidance of view-consistency, and generates representation with strong separability by a tensor nuclear norm regularizer to reveal the potential data cluster membership.
6. **IKMKC (Liu et al., 2020).** This algorithm incorporates kernel completion into clustering procedure to get rid of the limitation of pre-specified base kernels, and explicitly encourages incomplete kernels to reciprocally pad each other so as to better serve clustering.
7. **IMVTSC-MVI (Wen et al., 2021b).** This algorithm couples graph-learning manifold space and view-inferring feature space to exploit the hidden information of incomplete views, and forms intra-view representation by leveraging high-order correlations between views through a low-rank tensor regularizer.
8. **CPM-Nets (Zhang et al., 2019).** This algorithm gathers structured representation hidden into all views and samples via encoding networks instead of completing missing views based on the combination of available views, and handles patterns flexibly by improving the representation separability using a clustering-like classification loss.
9. **LSIMVC (Liu et al., 2022a).** This algorithm unifies graph embedding learning and representation learning into a framework based on matrix factorization, and extracts sparse individual representation and consensus representation as well as structured representation via $\ell_1$ norm and local graph embedding constraints.
10. **GSRIMC (Li et al., 2022a).** This algorithm avoids complicated partial feature recovering by just exploiting the pre-computed sub-graphs belonging to each view to eliminate mistakes in optimization, and decouples the polished graph structure from biased error via a tensor nuclear paradigm.
11. **COMPLETER (Lin et al., 2021).** This algorithm performs across-view sample restoration and representation learning alternatively in perspective of information theory, and maximizes the mutual information and minimizes the conditional entropy via contrastive learning and dual prediction respectively to aggregate informative features.

12. **TCIMC (Xia et al., 2022a).** This algorithm leverages the spatial structure and complementary representation through Schatten $p$-norm to refine the similarity of interview graph, and encodes the low-rank information embedded into unmissing data via connected component constraint to appropriately determine the number of clusters.

13. **LRGR-IMVC (Cui et al., 2022).** This algorithm employs graph embedding and Laplacian rank constraints to reduce the negative influence of incomplete samples on latent co-correlations learned across views, and concurrently utilizes local, global and potential representation within views to adaptively restore partial structure.

14. **BGIMVSC (Sun et al., 2023).** This algorithm produces probability consistent characteristics with all true elements by reformulating the standard spectral clustering procedure, and adopts a weighted learning mechanism to automatically measure the contributions of diverse views in the framework optimization.

15. **NGSP-CGL (Wong et al., 2023).** This algorithm characterizes the data distribution by taking advantage of both pair-wise neighbor and group-wise structure features, and forms clustering-friendly representation by improving the quality of unified graph matrix using a neighbor-group embedding item.

16. **PIMVC (Deng et al., 2023).** This algorithm utilizes the projection learning strategy to alleviate the impact of information imbalance between views, and exploits heterogeneous geometric features inside the samples by imposing an uncorrelated scatter matrix constraint and a graph constraint on the union subspace.

17. **ProImp (Li et al., 2023a).** This algorithm applies a dual-stream network framework containing a twin contrastive loss and a twin attention layer to support the within-cluster instance commonality, and recovers data by utilizing the sample relationship from observed views to retain the cross-view instance versatility.

18. **HCP-IMSC (Li et al., 2022b).** This algorithm maintains high-order view and sample correlations by constructing several three-order tensor affinities to recover the space structure of partial instances, and merges view-wise similarities in a self-weighted way to restrict samples with incomplete views to their neighborhoods.

19. **APADC (Xu et al., 2023).** This algorithm relieves the impact of inaccurate imputed values by adaptively projecting features that are learned by autoencoders into a shared space, and aligns the distribution between incomplete and complete data by minimizing the mean discrepancy loss to facilitate common representation.

20. **HCLS-CGL (Wen et al., 2023).** This algorithm designs a confidence graph with set-neighbor affinity to characterize the neighbor probability between any two instances, and captures consistent representation at graph structure level instead of from original data to reduce the impact of noise and outliers.

## J. More Observations and Analysis from Table 1 and Table 2

1. ToRES wins the ability to produce eye-catching clustering results. For instance, ToRES defeats all comparison methods in ACC under PID = 20% and 40%; ToRES consistently makes the best results on AwAfea in ACC, NMI and Purity; ToRES outperforms all competitors with remarkable margins on all datasets under PID = 40% in Purity. Under other circumstances, ToRES still can provide competitive results. This well illustrates that ToRES is effective for diverse IMVC problems.

2. ToRES can properly deal with all of the early-mentioned datasets. By contrast, some of comparison algorithms can not run on certain datasets and report the "\" error due to the considerable memory overheads or the limitations of algorithm itself. Especially on EMNIST, which consists of 280000 samples, all competitors output out-of-memory errors while ToRES still can normally execute. This demonstrates that ToRES is memory-friendly and more practical for coping with IMVC tasks.

3. ToRES can make clear improvements compared to the second best algorithms. For example, on Webkb with PID = 40% and in terms of ACC, NMI and Purity, it exceeds APMC, HCP-IMSC and APMC by 4.95%, 12.94% and 4.95%, respectively. On Wikifea with PID = 20%, it exceeds HCLS-CGL, IMVTSC-MVI and HCLS-CGL by 2.84%, 2.55% and 0.66%, respectively. Although receiving some sub-optimal results under certain situations, they are not significantly lower than the best results and still remain competitive. This indicates that ToRES is capable of producing comparable results in multiple situations.

4. ToRES consistently outperforms methods IMSC-AGL, IMG, TMBSD, IKMKC, LSIMVC, BGIMVSC, APADC, etc, which is mainly due to the facts that it can deeply mine multi-view representation by jointly utilizing two types of prototypes, and that it can well preserve the diversity information of original samples by directly learning discrete cluster labels from incomplete data. Comparison algorithms like IMSC-AGL, IMG and LSIMVC are with at least three hyper-parameters, and however even with the assistance of so many hyper-parameters, they still can not defeat us in terms of any one metric, which once again highlights the superiorities of our proposed ToRES.

## K. Additional Ablation Studies

### K.1. Ablation for Optimizable Prototypes

Besides being able to generate prototypes through learning, we also can utilize heuristic sampling schemes to generate prototypes. The difference is that the prototype generation process of heuristic schemes is finished before optimizing other variables. That is, the prototype generation process of heuristic schemes and the subsequent variable optimization process are separated from each other. By contrast, our optimizable prototype strategy enables prototypes and other variables to communicate so as to move towards a mutually reinforcing direction. To highlight its advantages, we conduct four group comparison experiments, as shown in Fig. (5), where "EmEF", "EmF" and "EF" represent the clustering results acquired based on finished $\mathbf{E}_m$ and $\mathbf{E}$ (i.e., finished view-wise prototypes and finished cross-view prototypes), finished $\mathbf{E}_m$ (i.e., finished view-wise prototypes), finished $\mathbf{E}$ (i.e., finished cross-wise prototypes), respectively. "Ours" is the results based on optimizable view-wise prototypes and optimizable cross-view prototypes. As seen, we receive the best clustering results in most cases, which demonstrates that the optimizable prototypes are more desirable.

### K.2. Ablation for the Strategy of Directly Learning Discrete Cluster Labels

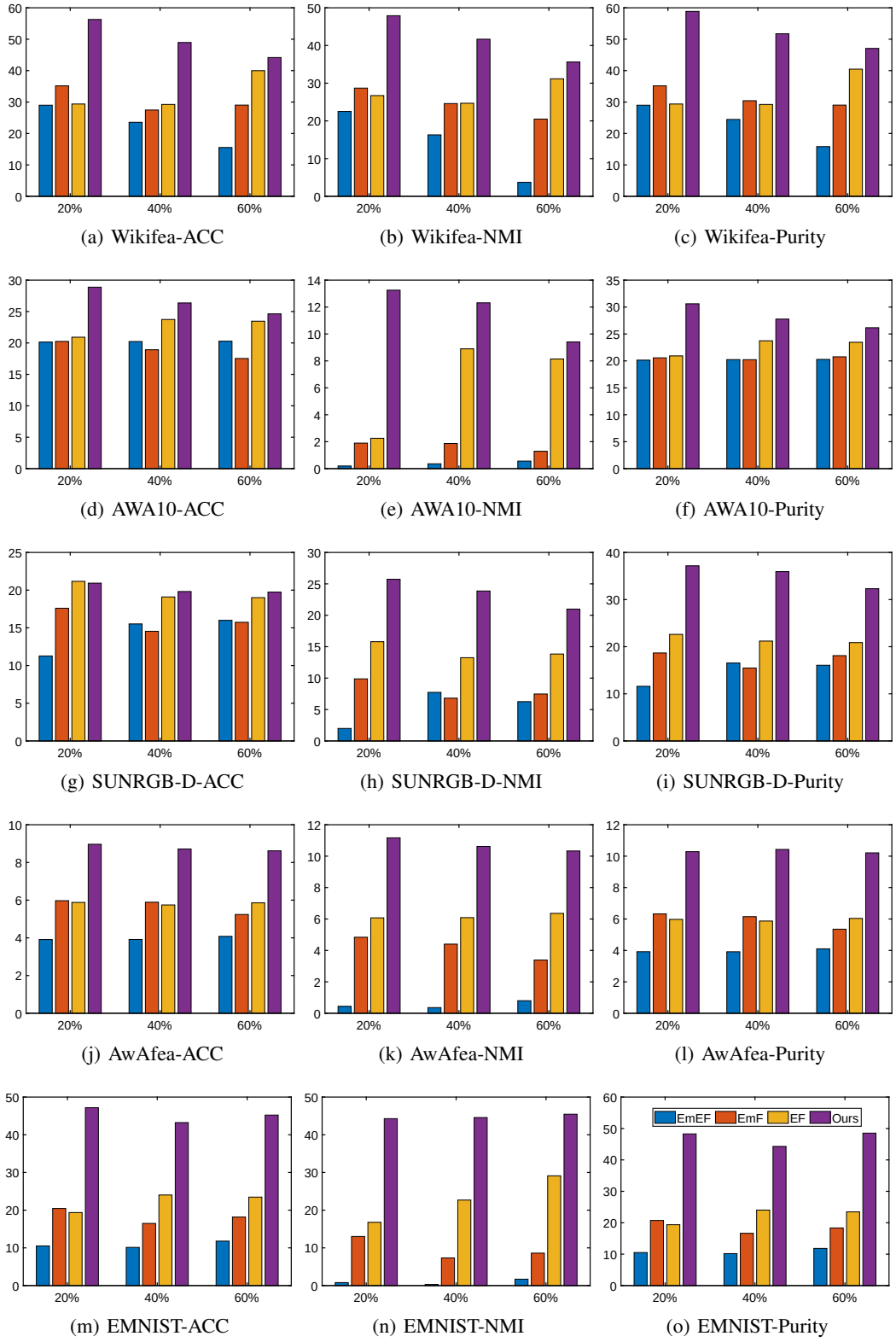
Current IMVC methods usually adopt a two-step strategy to generate the cluster labels, that is, first forming the spectral embedding and then executing the spectral partitioning operation on it. We hold that this two-step strategy is not very recommendable since the final labels are heavily dependent on the quality of formed spectral embedding. Unlike them, we directly optimize the discrete labels from incomplete data. This allows the final results to directly come from the original samples, not only well maintaining the diversity information of multi-view data, but also making the clustering procedure and variable optimization procedure able to negotiate with each other to enhance mutually. To verify this point, we carry out some ablation studies, and present the results in Table 6. It can be seen that in addition to enjoying zero-variance, the mean of our clustering results is higher in most cases. Also, our time consumption is consistently lower than that of two-step strategy. These clearly illustrate that the strategy of directly learning discrete cluster labels from incomplete data is effective and preferable, and that our ToRES is with the ability to generate both stable and superior clustering results.

*Table 6.* Clustering Result Comparison Between Two-step Strategy and Ours.

| Dataset | Ablation Study | 20% | | | | 40% | | | | 60% | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ACC | NMI | Purity | Time | ACC | NMI | Purity | Time | ACC | NMI | Purity | Time |
| WebKB | Two-step | 78.50±0.00 | 18.37±0.00 | 78.50±0.00 | 11.68 | 71.55±0.00 | 7.92±0.00 | 78.12±0.00 | 11.78 | 70.50±0.00 | **6.78±0.00** | 78.12±0.00 | 11.83 |
| | Ours | **86.20±0.00** | **32.61±0.00** | **86.20±0.00** | **3.37** | **85.35±0.00** | **31.95±0.00** | **85.35±0.00** | **3.43** | **73.83±0.00** | 4.11±0.00 | **78.12±0.00** | **3.39** |
| Wiki | Two-step | 52.75±1.61 | 45.46±0.88 | 56.90±1.42 | 53.30 | 48.60±2.60 | 38.75±1.76 | **53.31±2.39** | 54.72 | 43.74±2.19 | 31.15±0.93 | **47.86±1.85** | 46.93 |
| | Ours | **56.28±0.00** | **47.90±0.00** | **58.86±0.00** | **1.36** | **48.95±0.00** | **41.67±0.00** | 51.74±0.00 | **1.36** | **44.17±0.00** | **35.63±0.00** | 47.07±0.00 | **1.44** |
| AWA10 | Two-step | 25.97±1.03 | 10.44±0.35 | 29.09±0.58 | 66.22 | 24.27±0.92 | 9.72±0.47 | **28.02±0.76** | 71.43 | 21.93±0.80 | 9.27±0.35 | 25.41±0.50 | 71.69 |
| | Ours | **28.88±0.00** | **13.25±0.00** | **30.60±0.00** | **17.57** | **26.37±0.00** | **12.31±0.00** | 27.76±0.00 | **29.19** | **24.63±0.00** | **9.41±0.00** | **26.16±0.00** | **31.86** |
| SUNRGBD | Two-step | 17.47±0.56 | 22.30±0.24 | 35.86±0.41 | 516.95 | 16.77±0.40 | 19.85±0.21 | 33.04±0.31 | 481.63 | 16.96±0.42 | 18.38±0.21 | 31.51±0.32 | 511.87 |
| | Ours | **20.93±0.00** | **25.73±0.00** | **37.16±0.00** | **46.13** | **19.82±0.00** | **23.87±0.00** | **35.94±0.00** | **76.13** | **19.75±0.00** | **20.98±0.00** | **32.31±0.00** | **76.62** |
| AwAfea | Two-step | 8.94±0.14 | 10.09±0.19 | **10.39±0.19** | 803.32 | **8.85±0.08** | 10.01±0.11 | 11.06±0.10 | 768.91 | 8.61±0.27 | 9.54±0.26 | **10.50±0.28** | 728.40 |
| | Ours | **8.96±0.00** | **11.17±0.00** | 10.29±0.00 | **238.92** | 8.72±0.00 | **10.62±0.00** | **10.43±0.00** | **230.98** | **8.62±0.00** | **10.33±0.00** | 10.21±0.00 | **234.21** |

## L. Single-view Effectiveness and Performance Comparison

Sometimes we may encounter some datasets with only one view. Previously, we mainly concentrate on multi-view datasets. When facing single-view data, how is the performance of the algorithm? To this end, we conduct clustering on one view of the early-mentioned dataset. Table 7 reports the single-view clustering results. It is easy to find that methods APMC, IMG, IMVTSC-MVI, CPM-Nets, GSRIMC, COMPLETER, ProImp and APADC are incapable of coping with single-view data. Although IMSC-AGL, AWP, TMBSD, IKMKC, LSIMVC, TCIMC, LRGR-IMVC, NGSP-CGL and HCP-IMSC can normally execute, they typically generate sub-optimal results. By comparison, in addition to running properly on single-view




(a) Wikifea-ACC (b) Wikifea-NMI (c) Wikifea-Purity

(d) AWA10-ACC (e) AWA10-NMI (f) AWA10-Purity

(g) SUNRGB-D-ACC (h) SUNRGB-D-NMI (i) SUNRGB-D-Purity

(j) AwAfea-ACC (k) AwAfea-NMI (l) AwAfea-Purity

(m) EMNIST-ACC (n) EMNIST-NMI (o) EMNIST-Purity

*Figure 5.* Ablation study for optimizable $\mathbf{E}_m$ and $\mathbf{E}$.

data, we are also able to outperform the strong competitors in most cases. These results well reveal that our ToRES not only is applicable for handling single-view data, but also can provide relatively superior clustering results.

*Table 7.* Clustering Result Comparison on Single-view Data.

| Methods | 20% | | | | | 40% | | | | | 60% | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | WebKB | Wikifea | AWA-10 | SUNRGB-D | AwAfea | WebKB | Wikifea | AWA-10 | SUNRGB-D | AwAfea | WebKB | Wikifea | AWA-10 | SUNRGB-D | AwAfea |
| IMSC-AGL | 73.93±0.00 | 26.47±1.04 | 16.31±0.28 | 7.22±0.23 | \ | 71.46±0.00 | 22.07±1.19 | 15.78±0.66 | 6.78±0.27 | \ | 68.32±0.00 | 20.83±1.65 | 15.78±0.66 | 6.30±0.19 | \ |
| AWP | 69.17±0.00 | 51.88±0.00 | 18.99±0.00 | 6.90±0.00 | 3.41±0.00 | 73.74±0.00 | 48.39±0.00 | 17.58±0.00 | 9.58±0.00 | 4.02±0.00 | 70.78±0.00 | 27.50±1.02 | 14.23±0.00 | 5.86±0.00 | 5.13±0.00 |
| APMC | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| IMG | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| TMBSD | 61.77±0.85 | 39.84±0.61 | 13.65±0.07 | 7.28±0.09 | \ | 61.23±0.13 | 33.59±0.79 | 13.36±0.01 | 7.93±0.06 | \ | 58.15±0.00 | 28.34±0.99 | 13.36±0.01 | 9.04±0.02 | \ |
| IKMKC | 72.31±0.00 | 52.69±0.81 | 19.31±0.53 | 10.23±0.25 | \ | 68.51±0.00 | 50.27±1.32 | 17.42±0.02 | 9.70±0.30 | \ | 68.51±0.00 | 45.12±0.02 | 16.74±0.27 | 9.35±0.26 | \ |
| IMVTSC-MVI | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| CPM-Nets | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| LSIMVC | 74.50±0.00 | 36.11±0.00 | 18.38±0.07 | 9.99±0.25 | \ | 68.41±0.00 | 33.81±0.00 | 17.87±0.08 | 10.21±0.20 | \ | 64.03±0.00 | 30.57±0.00 | 18.44±0.03 | 10.34±0.25 | \ |
| GSRIMC | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| COMPLETER | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| TCIMC | 78.02±0.00 | 12.82±0.36 | 19.16±0.02 | \ | \ | 78.02±0.00 | 13.53±0.36 | 15.09±0.21 | \ | \ | 78.02±0.00 | 13.24±0.23 | 18.17±0.01 | \ | \ |
| LRGR-IMVC | 77.64±0.00 | 52.11±0.15 | 19.32±0.16 | 10.67±0.16 | \ | 77.64±0.00 | 47.49±0.00 | 18.53±0.05 | 10.95±0.19 | \ | 77.64±0.00 | 39.01±0.01 | 18.04±0.03 | 10.91±0.11 | \ |
| BGIMVSC | 77.93±0.00 | 37.44±1.00 | 19.15±0.05 | 11.51±0.36 | \ | 77.93±0.00 | 35.39±1.22 | 18.75±0.05 | **11.44±0.70** | \ | 77.74±0.00 | 32.39±1.24 | 18.11±0.07 | **11.23±0.45** | \ |
| NGSP-CGL | 77.16±0.00 | 37.89±2.38 | 18.84±0.40 | 10.89±0.23 | 6.26±0.06 | 76.88±0.00 | 34.23±1.94 | 18.28±0.31 | 11.25±0.22 | 5.96±0.05 | 76.88±0.00 | 32.79±1.66 | 17.79±0.36 | 10.66±0.65 | 5.72±0.06 |
| PIMVC | 76.74±1.56 | 54.39±0.03 | 19.90±0.06 | 10.26±0.40 | \ | 76.23±1.23 | **51.12±0.00** | 18.18±0.03 | 10.16±0.20 | \ | 75.21±1.02 | 45.54±0.02 | 18.56±0.42 | 10.50±0.28 | \ |
| ProImp | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| HCP-IMSC | 78.02±0.00 | 49.28±0.04 | 17.40±0.10 | 11.14±0.27 | \ | 78.02±0.00 | 47.33±0.98 | 16.63±0.03 | 10.97±0.29 | \ | 78.02±0.00 | 41.97±0.70 | 17.86±0.14 | 11.05±0.27 | \ |
| APADC | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ | \ |
| HCLS-CGL | 77.16±0.00 | 35.13±0.05 | 19.28±0.00 | **11.59±0.28** | \ | 76.97±0.00 | 29.10±0.00 | 18.59±0.00 | 11.33±0.13 | \ | 79.59±0.00 | 28.87±0.02 | 18.33±0.00 | 10.98±0.21 | \ |
| Ours | **80.69±0.00** | **55.13±0.00** | **19.97±0.00** | 10.45±0.00 | **6.63±0.00** | **80.11±0.00** | 48.43±0.00 | **18.78±0.00** | 10.91±0.00 | **6.24±0.00** | **79.64±0.00** | **46.16±0.00** | **18.89±0.00** | 10.94±0.00 | **5.79±0.00** |

## M. Effectiveness in Extracting Multi-view Information

Multi-view datasets typically contain more diverse information than single-view datasets. To demonstrate that ToRES can effectively extract the information from multiple views for better clustering, we cluster each view separately, and compare the results of each individual view with the results obtained from multiple views, as presented in Table 8, where "V1" represents the clustering results on view 1, and "MV" represents the results acquired from multiple views. Similar meanings are for "V2" $\sim$ "V6". From Table 8, we can observe that the generated multi-view results are superior to single-view results in most cases. This well illustrates the effectiveness of the proposed ToRES in extracting information from multiple views.

*Table 8.* Effectiveness of ToRES in Extracting Multi-view Information.

| Dataset | Webkb | | | Wikifea | | | AWA10 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | V1 | V2 | MV | V1 | V2 | MV | V1 | V2 | V3 | V4 | V5 | V6 | MV |
| 20% | 62.51 | 80.69 | **86.20** | 17.10 | 55.13 | **56.28** | 18.21 | 19.97 | 19.09 | 21.60 | 20.92 | 23.63 | **28.88** |
| 40% | 61.18 | 80.11 | **85.35** | 16.15 | 48.43 | **48.95** | 18.06 | 18.78 | 19.07 | 20.16 | 19.66 | 25.73 | **26.37** |
| 60% | 57.09 | **79.64** | 73.83 | 16.05 | **46.16** | 44.17 | 18.64 | 18.89 | 20.67 | 20.66 | 19.93 | 23.15 | **24.63** |
| | SUNRGB-D | | | EMNIST | | | AwAfea | | | | | | |
| | V1 | V2 | MV | V1 | V2 | MV | V1 | V2 | V3 | V4 | V5 | V6 | MV |
| 20% | 17.03 | 10.45 | **20.93** | 35.41 | 42.80 | **47.18** | 6.34 | 6.63 | 6.61 | 7.79 | 6.74 | 8.26 | **8.96** |
| 40% | 16.65 | 10.91 | **19.82** | 35.85 | 39.71 | **43.23** | 6.15 | 6.24 | 6.46 | 7.26 | 6.27 | 7.81 | **8.72** |
| 60% | 16.07 | 10.94 | **19.75** | 31.54 | 35.89 | **45.22** | 5.82 | 5.79 | 5.98 | 6.82 | 6.31 | 7.44 | **8.62** |

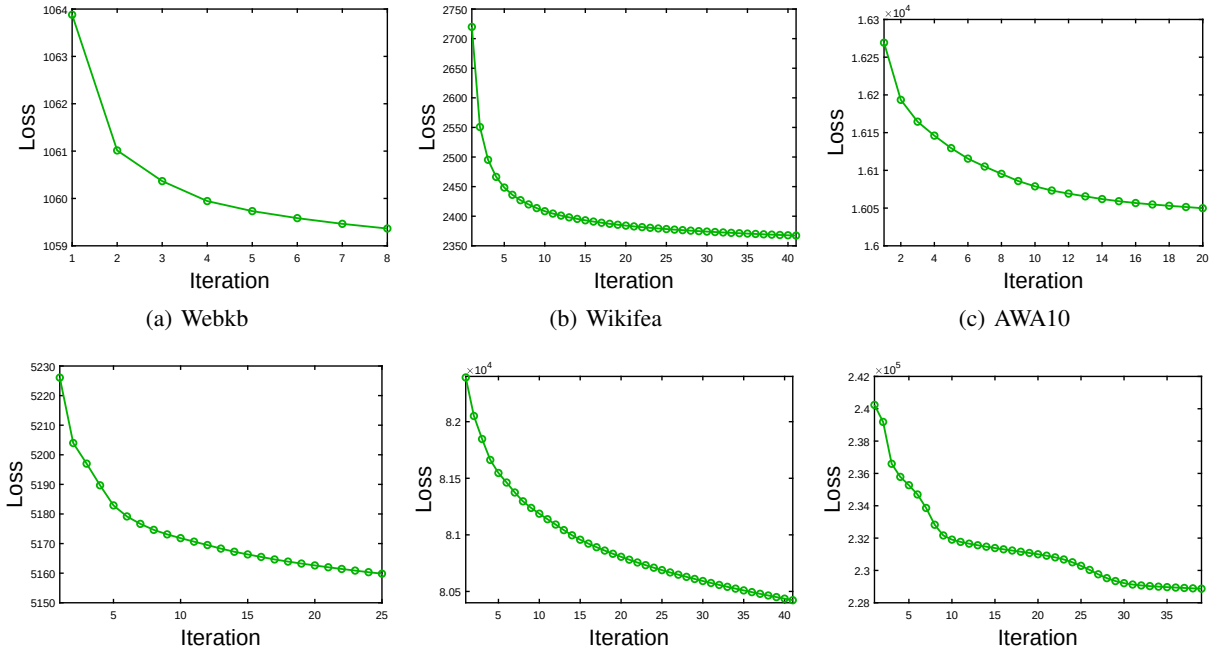
## N. Convergence

In this section, we conduct convergence analysis on the proposed ToRES from theoretical and experimental perspectives respectively.

Denote the objective of Eq. (3) as $\mathcal{F}\left(\{\mathbf{G}_m\}_{m=1}^M, \mathbf{E}, \mathbf{O}, \mathbf{L}, \{\mathbf{E}_m\}_{m=1}^M\right)$, the solutions at the $k$-th iteration as $\{\mathbf{G}_m^{(k)}\}_{m=1}^M$, $\mathbf{E}^{(k)}$, $\mathbf{O}^{(k)}$, $\mathbf{L}^{(k)}$, $\{\mathbf{E}_m^{(k)}\}_{m=1}^M$. When optimizing $\mathbf{O}$, based on the given $\{\mathbf{G}_m^{(k)}\}_{m=1}^M$, $\mathbf{E}^{(k)}$, $\mathbf{L}^{(k)}$ and $\{\mathbf{E}_m^{(k)}\}_{m=1}^M$, the solution $\mathbf{O}^{(k+1)}$ at the $(k+1)$-th iteration can be obtained by Algorithm 1 or Algorithm 2. Thus, we always have

$$\mathcal{F}\left(\{\mathbf{G}_m^{(k)}\}_{m=1}^M, \mathbf{E}^{(k)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right) \leq \mathcal{F}\left(\{\mathbf{G}_m^{(k)}\}_{m=1}^M, \mathbf{E}^{(k)}, \mathbf{O}^{(k)}, \mathbf{L}^{(k)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right). \tag{62}$$

When optimizing $\mathbf{L}$, based on the given $\{\mathbf{G}_m^{(k)}\}_{m=1}^M$, $\mathbf{E}^{(k)}$, $\mathbf{O}^{(k+1)}$ and $\{\mathbf{E}_m^{(k)}\}_{m=1}^M$, the solution $\mathbf{L}^{(k+1)}$ at the $(k+1)$-th




(a) Webkb (b) Wikifea (c) AWA10

(d) SUNRGB-D (e) AwAfea (f) EMNIST

*Figure 6.* Convergence Curves.

iteration can be obtained by Eq. (14). Thus, we have

$$\mathcal{F}\left(\{\mathbf{G}_m^{(k)}\}_{m=1}^M, \mathbf{E}^{(k)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k+1)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right) \leq \mathcal{F}\left(\{\mathbf{G}_m^{(k)}\}_{m=1}^M, \mathbf{E}^{(k)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right). \tag{63}$$

When optimizing each $\mathbf{G}_m$, based on the given $\mathbf{E}^{(k)}$, $\mathbf{O}^{(k+1)}$, $\mathbf{L}^{(k+1)}$ and $\{\mathbf{E}_m^{(k)}\}_{m=1}^M$, the solution $\mathbf{G}_m^{(k+1)}$ at the $(k+1)$-th iteration can be acquired by Eq. (16). Thus, for all $\{\mathbf{G}_m^{(k+1)}\}_{m=1}$, we have

$$\mathcal{F}\left(\{\mathbf{G}_m^{(k+1)}\}_{m=1}^M, \mathbf{E}^{(k)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k+1)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right) \leq \mathcal{F}\left(\{\mathbf{G}_m^{(k)}\}_{m=1}^M, \mathbf{E}^{(k)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k+1)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right). \tag{64}$$

When optimizing $\mathbf{E}$, based on the given $\{\mathbf{G}_m^{(k+1)}\}_{m=1}^M$, $\mathbf{O}^{(k+1)}$, $\mathbf{L}^{(k+1)}$ and $\{\mathbf{E}_m^{(k)}\}_{m=1}^M$, the solution $\mathbf{E}^{(k+1)}$ at the $(k+1)$-th iteration can be acquired by Eq. (19). Thus, we have

$$\mathcal{F}\left(\{\mathbf{G}_m^{(k+1)}\}_{m=1}^M, \mathbf{E}^{(k+1)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k+1)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right) \leq \mathcal{F}\left(\{\mathbf{G}_m^{(k+1)}\}_{m=1}^M, \mathbf{E}^{(k)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k+1)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right). \tag{65}$$

When optimizing each $\mathbf{E}_m$, based on the given $\{\mathbf{G}_m^{(k+1)}\}_{m=1}^M$, $\mathbf{E}^{(k+1)}$, $\mathbf{O}^{(k+1)}$ and $\mathbf{L}^{(k+1)}$, the solution $\mathbf{E}_m^{(k+1)}$ at the $(k+1)$-th iteration can be acquired by Eq. (21). Thus, for all $\{\mathbf{E}_m^{(k+1)}\}_{m=1}^M$, we have

$$\mathcal{F}\left(\{\mathbf{G}_m^{(k+1)}\}_{m=1}^M, \mathbf{E}^{(k+1)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k+1)}, \{\mathbf{E}_m^{(k+1)}\}_{m=1}^M\right) \leq \mathcal{F}\left(\{\mathbf{G}_m^{(k+1)}\}_{m=1}^M, \mathbf{E}^{(k+1)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k+1)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right). \tag{66}$$

Combined with Eqs. (62), (63), (64), (65) and (66), we can get that

$$\mathcal{F}\left(\{\mathbf{G}_m^{(k+1)}\}_{m=1}^M, \mathbf{E}^{(k+1)}, \mathbf{O}^{(k+1)}, \mathbf{L}^{(k+1)}, \{\mathbf{E}_m^{(k+1)}\}_{m=1}^M\right) \leq \mathcal{F}\left(\{\mathbf{G}_m^{(k)}\}_{m=1}^M, \mathbf{E}^{(k)}, \mathbf{O}^{(k)}, \mathbf{L}^{(k)}, \{\mathbf{E}_m^{(k)}\}_{m=1}^M\right), \tag{67}$$

which indicates that during the process of alternatively optimizing $\{\mathbf{G}_m\}_{m=1}^M, \mathbf{E}, \mathbf{O}, \mathbf{L}$ and $\{\mathbf{E}_m\}_{m=1}^M$, the objective in Eq. (3) is monotonically decreasing. In addition, the objective has the lower bound. According to alternating optimization theory, our ToRES will be convergent.

Then, to validate its convergence experimentally, we draw the evolution curve of the objective during iterations, as shown in Fig. (6). It can be seen that the objective value decreases monotonically, and reaches a stable state within about forty iterations, which well demonstrates the convergence of ToRES.