



High-order Complementarity Induced Fast Multi-View Clustering with Enhanced Tensor Rank Minimization

Jintian Ji

Key Laboratory of Big Data & Artificial Intelligence in Transportation, Ministry of Education, Beijing Jiaotong University, Beijing, China

School of Computer and Information Technology, Beijing Jiaotong University,
Beijing, China
22120385@bjtu.edu.cn

Songhe Feng*

Key Laboratory of Big Data & Artificial Intelligence in Transportation, Ministry of Education, Beijing Jiaotong University, Beijing 100044, China

School of Computer and Information Technology, Beijing Jiaotong University,
Beijing, China
shfeng@bjtu.edu.cn

ABSTRACT

Recently, tensor-based multi-view clustering methods have achieved promising results, primarily benefited from their superior ability in exploring high-order consistent information among views. Despite significant progress, these methods inevitably suffer from several drawbacks: 1) Extremely high computational complexity restricts their feasibility for large-scale data sets. 2) Prevalently adopted tensor rank approximations (e.g., Tensor Nuclear Norm (TNN)) tend to under-penalize small singular values, resulting in noise residuals. 3) Tensor structure is rarely utilized for high-order complementarity investigation. In light of this, we propose High-order Complementarity Induced Fast Multi-View Clustering with Enhanced Tensor Rank Minimization (**CFMVC-ETR**). Specifically, two sets of representation matrices are learned from original multi-view data via the matrix factorization mechanism with a group of base matrices, which are further reconstructed into the consistent tensor and the complementary tensor, respectively. Subsequently, a novel Enhanced Tensor Rank is imposed on the consistent tensor, which is a tighter approximation of the tensor rank and is more noisy-robust to explore the high-order consistency. Meanwhile, a tensor-level constraint termed Tensorial Exclusive Regularization is proposed on the complementary tensor to enhance the view-specific feature and well capture the high-order complementarity. Moreover, we adopt a concatenation-fusion approach to integrate these two parts, deriving a discriminative unified embedding for the clustering task. We solve CFMVC-ETR by an efficient algorithm with good convergence. Extensive experiments on nine challenging data sets demonstrate the superiority of the proposed method.

CCS CONCEPTS

- Computing methodologies → Cluster analysis.

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MM '23, October 29–November 3, 2023, Ottawa, ON, Canada.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-0108-5/23/10...\$15.00
<https://doi.org/10.1145/3581783.3611733>

KEYWORDS

Tensor-based multi-view clustering; Matrix factorization; High-order consistency; High-order complementarity

ACM Reference Format:

Jintian Ji and Songhe Feng. 2023. High-order Complementarity Induced Fast Multi-View Clustering with Enhanced Tensor Rank Minimization. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29–November 3, 2023, Ottawa, ON, Canada*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3581783.3611733>

1 INTRODUCTION

Multi-view data refers to data that represents an object from multiple perspectives or dimensions. With the increasing improvement of data technology, such multi-view data can be easily obtained, for example, in autonomous driving, different objects can be captured by infrared sensors, lidar sensors, radar sensors, etc., which depict different attributes of the same object. So multi-view data contains more comprehensive information than single-view data. Multi-view clustering aims to explore the consistent and complementary information in multi-view data to improve the performance of traditional clustering such as spectral clustering [31]. These approaches allow for a more comprehensive understanding of the underlying structure of the data and can lead to more accurate clustering results.

The existing multi-view clustering methods can be broadly categorized into two types based on whether high-order correlation among multi-view representations is explored: tensor-based methods and non-tensor methods. Non-tensor methods [12, 27, 32, 37, 40, 45, 46] employ matrix-level constraints to explore the point-to-point relationship within one view or between pairs of views. For instance, [3] diversifies the clustering structure using the Hilbert Schmidt Independence Criterion (HSIC) on pairs of affinity matrices. The graph Laplacian regularization is employed in [12] to retain the graph geometric features in each view. Huang *et al.* [13] separately explore the consistent information and complementary information by dividing the affinity matrix in each view into two parts. However, non-tensor methods are limited to processing and constraining all views simultaneously, leading to unsatisfactory clustering results. So, the tensor-based methods [4, 5, 9, 28, 35, 39, 48] are born to exploit the high-order correlation among views, where affinity matrices in all views are merged into a representation tensor, and tensor-level constraints are imposed to capture the view-to-view relationship. For example, [43] first extends the low-rank constraint

from the matrix level to the tensor level by minimizing the linear combination of the nuclear norm of all views to exploit the high-order correlation among views. [22, 23, 48] propose a tensor average rank to represent the true rank of the tensor and prove that the Tensor Nuclear Norm (TNN) based on the tensor Singular Value Decomposition (t-SVD) is the tightest convex approximation to the tensor average rank. [39] adopts TNN on the rotated tensor to capture the high-order consistency. However, TNN is a loose and biased approximation of tensor rank, which treats different singular values equally, leading to noise residuals and loss of important information. Therefore, several non-convex approximations, such as Tensor Logarithmic Schatten-p Norm (TLS_pN) [11] and Tensorial Arctangent Rank (TAR) [26], have been proposed to enhance the characterization of tensor low-rankness.

Although existing methods have obtained promising achievements, they inevitably suffer from three key challenges: (1) The tensor-based operations like t-SVD and Fast Fourier Transformation (FFT) [6] usually take extremely high computational complexity, which limits their applicability to handle large-scale data sets. (2) Existing tensor rank approximation methods, such as TNN and TLS_pN , fail to penalize small singular values sufficiently, resulting in noisy residuals that disrupt the clustering structure. (3) As samples in similar classes share semblable clustering structures across views, it is challenging for existing tensor-based methods to distinguish them by focusing on high-order consistency learning only.

To tackle the aforementioned problems, in this paper, we propose a fast multi-view clustering framework termed **CFMVC-ETR**, which integrates the matrix factorization technique, tensor-level constraints, complementarity learning, and consistency learning into a unified optimization model. Specifically, we employ the matrix factorization technique to learn two groups of low-dimension representation matrices from the original multi-view features, which are further reconstructed into the consistent tensor and the complementary tensor, respectively. In order to learn the high-order consistency, the consistent tensor is constrained by the minimization of a noisy-robust Enhanced Tensor Rank (ETR). Additionally, we introduce a novel Tensorial Exclusive Regularization (TER) on the complementary tensor to discover the high-order complementarity across views. Finally, we adopt a concatenation-fusion strategy to integrate these two parts into a discriminative unified representation for the clustering task. In summary, we conclude the novelty and main contributions of our model as follows:

(1) We employ the matrix decomposition mechanism to construct two representation tensors, namely the consistent tensor and the complementary tensor, which enjoy a space-saving size that reduces the complexity of tensor-related operations, allowing our model to efficiently handle large data sets.

(2) To better explore the high-order consistency, we introduce a novel method called Enhanced Tensor Rank (ETR) to approximate the rank of the consistent tensor, which is more noisy-robust than existing methods like TNN. Moreover, we design a Tensorial Exclusive Regularization (TER) for the complementary tensor to enhance the view-specific feature and capture the high-order complementary information.

(3) We adopt a concatenation-fusion strategy to fuse the learned consistent tensor and complementary tensor into a discriminative

unified embedding for the clustering task, which reveals a distinct clustering structure of samples, especially for samples belonging to similar classes.

(4) An efficient algorithm is proposed to solve our optimization problem, which enjoys economical computational complexity and mathematically proven convergence. Extensive experimental results on nine challenging data sets demonstrate the superiority of our model.

2 PRELIMINARY

2.1 Notations

In this paper, bold-lowercase \mathbf{x} and lowercase x represent a vector and a scalar, respectively. Bold-uppercase \mathbf{X} denotes the matrix. \mathbf{I}_n is the $n \times n$ identity matrix. Calligraphy letter denotes the tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. \mathcal{X}^k is the k -th frontal slice of \mathcal{X} . \mathcal{X}_f means the Fast Fourier Transformation (FFT) along the third dimension of tensor \mathcal{X} . Multi-view data set with n samples and m views is denoted as $\{\mathbf{X}^v\}_{v=1}^m$, where $\mathbf{X}^v \in \mathbb{R}^{n \times d^v}$, d^v is the feature dimension of v -th view. c is the cluster number of this data set.

2.2 Single-view matrix factorization clustering

The matrix factorization technique [2, 7, 24] is an effective method for large-scale data clustering, especially for high-dimensional data sets. It aims to decompose the original feature $\mathbf{X} \in \mathbb{R}^{n \times d}$ into two parts, denoted as representation matrix $\mathbf{Z} \in \mathbb{R}^{n \times c}$ and base matrix $\mathbf{H} \in \mathbb{R}^{c \times d}$. As one of the typical methods, Non-negative matrix factorization (NMF) [1, 7] regularizes \mathbf{Z} and \mathbf{H} to be non-negative, which is formulated as

$$\min_{\mathbf{Z} \geq 0, \mathbf{H} \geq 0} \mathcal{L}(\mathbf{X}, \mathbf{ZH}), \quad (1)$$

where $\mathcal{L}(\cdot)$ is the loss function that is usually measured by the Frobenius norm [19]. To improve the distinguishability of the representation matrix \mathbf{Z} , [8, 33] further impose orthogonality constraints on the base matrix or representation matrix.

After getting the low-dimension representation matrix \mathbf{Z} , the clustering results can be obtained efficiently by imposing traditional clustering methods such as k -means or spectral clustering algorithm.

2.3 Multi-view matrix factorization clustering

Multi-view matrix factorization aims to learn a consensus representation matrix \mathbf{Z} to reveal the clustering structure from multi-view data $\{\mathbf{X}^v\}_{v=1}^m$. We conclude the general formula of such methods as follows:

$$\begin{aligned} & \min_{\{\mathbf{Z}^v, \mathbf{H}^v\}_{v=1}^m} \mathcal{L}(\{\mathbf{X}^v, \mathbf{Z}^v \mathbf{H}^v\}_{v=1}^m) + \lambda \mathcal{R}(\{\mathbf{Z}^v\}_{v=1}^m) \\ & \text{s.t. } \forall v, \mathbf{H}^v \mathbf{H}^{vT} = \mathbf{I}_c, \end{aligned} \quad (2)$$

where λ is a trade-off parameter, $\{\mathbf{H}^v\}_{v=1}^m$ represents a group of base matrices, $\mathcal{R}(\cdot)$ denotes specific regularization terms on $\{\mathbf{Z}^v\}_{v=1}^m$. For example, [33] imposes orthogonality constraints to improve the discriminability of representation matrices. The graph Laplacian regularization is adopted in [47] to preserve the geometric structure. After learning the representation matrix \mathbf{Z}^v of each view, the consensus representation matrix \mathbf{Z} is obtained by fusion strategies (i.e., linear weighted fusion method [34] or spectral rotation fusion

method [44]), which is then fed into k -means or spectral clustering to acquire the final clustering results.

3 THE PROPOSED METHOD

Tensor-based multi-view clustering methods [9, 11, 39] aim to explore high-order correlation across views. Typically, these methods merge the affinity matrices into a tensor and apply a low-rank constraint to exploit high-order consistency. Such algorithms are generally formulated as follows:

$$\begin{aligned} & \min_{\{\mathbf{Z}^v, \mathbf{E}^v\}_{v=1}^m} \mathcal{T}(\mathcal{Z}) + \lambda_1 \mathcal{L}(\{\mathbf{E}^v\}) + \lambda_2 \mathcal{R}(\{\mathbf{Z}^v\}), \\ & \text{s.t. } \forall v, \{\mathbf{Z}^v, \mathbf{E}^v\} = \Theta(\mathbf{X}^v), \mathcal{Z} = \Phi(\mathbf{Z}^1, \dots, \mathbf{Z}^m), \end{aligned} \quad (3)$$

where $\Theta(\cdot)$ denotes the construction method of the affinity matrix $\mathbf{Z}^v \in \mathbb{R}^{n \times n}$ (e.g., predefined adjacency matrix method [38], self-representation method [39], etc.). $\mathcal{T}(\cdot)$ is the approximation of tensor rank. $\mathcal{L}(\cdot)$ is designed to constrain the error matrix \mathbf{E}^v , which usually adopts $\ell_{2,1}$ norm. $\mathcal{R}(\cdot)$ denotes some specific regularization such as graph Laplacian regularization [2]. $\Phi(\cdot)$ denotes the merging and rotating operation [39], which merges affinity matrices $\{\mathbf{Z}^v\}_{v=1}^m$ into a three-order tensor \mathcal{Z} with the dimension of $n \times m \times n$. λ_1 and λ_2 are two trade-off parameters.

Motivation 1: The representation tensor \mathcal{Z} generated by model (3) has a dimension of $n \times m \times n$, which leads to high computational complexity in tensor-related operations, such as Fast Fourier transformations (FFT), and makes it difficult to handle large-scale data sets. To tackle this problem, anchor-based methods [38, 41] select a subset of representative points covering the entire data set and construct affinity matrices between anchors and samples to reduce the size of the representation tensor. However, anchor-based methods have several limitations such as sub-optimal solutions [15] and the introduction of additional hyper-parameters leading to undesirable clustering performance. To overcome these limitations, we employ the matrix factorization mechanism (2) to construct the representation matrix of each view,

$$\forall v, \mathbf{X}^v = \mathbf{Z}^v \mathbf{H}^v + \mathbf{E}^v, \quad (4)$$

where $\mathbf{Z}^v \in \mathbb{R}^{n \times c}$. Here $\mathbf{E}^v \in \mathbb{R}^{n \times d^v}$ denotes reconstruction error. Accordingly, the new representation tensor $\mathcal{Z} = \Phi(\mathbf{Z}^1, \dots, \mathbf{Z}^m)$ enjoys a dimension of $c \times m \times n$, which can directly accelerate tensor-related operations without additional hyper-parameters.

Motivation 2: Generally speaking, the samples belonging to similar classes tend to exhibit consistent features across some views. Therefore, existing tensor-based methods often struggle to differentiate them only by exploring high-order consistency. To enhance the discriminability among similar classes, we consider both explorations of high-order consistency and high-order complementarity among views. Specifically, two types of representation matrices $\{\mathbf{Z}^v\}_{v=1}^m$ and $\{\mathbf{C}^v\}_{v=1}^m$ are obtained via

$$\forall v, \mathbf{X}^v = \mathbf{Z}^v \mathbf{H}^v + \mathbf{C}^v \mathbf{H}^v + \mathbf{E}^v, \quad (5)$$

where $\mathbf{Z}^v \in \mathbb{R}^{n \times c}$ and $\mathbf{C}^v \in \mathbb{R}^{n \times c}$. Then we merge $\{\mathbf{Z}^v\}_{v=1}^m$ and $\{\mathbf{C}^v\}_{v=1}^m$ into the consistent tensor $\mathcal{Z} \in \mathbb{R}^{c \times m \times n}$ and the complementary tensor $\mathbf{C} \in \mathbb{R}^{n \times c \times m}$, respectively. Moreover, we perform low-rank constraint on \mathcal{Z} to learn the high-order consistency among views. However, instead of adopting the TNN to approximate the tensor rank, we design a noise-robust Enhanced Tensor

Rank (ETR) which is defined in Definition 1. To discover high-order complementarity, we employ the proposed Tensorial Exclusive Regularisation (TER) represented in Definition 2 for the complementary tensor to capture the view-specific information.

Definition 1. Given a tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then the Enhanced Tensor Rank (ETR) is defined as:

$$\|\mathcal{Z}\|_{ETR} = \frac{1}{n_3} \sum_{k=1}^{n_3} \left\| \mathcal{Z}_f^k \right\|_{ETR} = \frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{i=1}^h \left(\frac{e^{\delta^2} S_f^k(i, i)}{\delta + S_f^k(i, i)} \right), \quad (6)$$

where $0 < \delta \leq 1$, $h = \min(n_1, n_2)$ and S_f is obtained by t-SVD of $\mathcal{Z}_f = \mathcal{U}_f \mathcal{S}_f \mathcal{V}_f^T$ in Fourier domain.

Definition 2. Given a tensor $C \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then the Tensorial Exclusive Regularisation (TER) is defined as:

$$\|C\|_{TER} = \sum_{v=1}^{n_3} \sum_{w \neq v} \langle C^v, C^w \rangle = \sum_{v=1}^{n_3} \sum_{w \neq v} \text{Tr}(C^v C^w)^T, \quad (7)$$

where $\langle \cdot, \cdot \rangle$ means inner product of two matrices.

By integrating Eq. (3), Eq. (5), Eq. (6), and Eq. (7), our CFMVC-ETR can be formulated as follows:

$$\begin{aligned} & \min_{\mathcal{Z}, C, E, \{\mathbf{H}^v\}_{v=1}^m} \|\mathcal{Z}\|_{ETR} + \alpha \|\mathbf{E}\|_{2,1} + \gamma \|C\|_{TER}, \\ & \text{s.t. } \forall v, \mathbf{X}^v = (\mathbf{Z}^v + \mathbf{C}^v) \mathbf{H}^v + \mathbf{E}^v, \mathbf{H}^v \mathbf{H}^{vT} = \mathbf{I}_c, \\ & \quad \mathbf{Z}^{vT} \mathbf{Z}^v = \mathbf{I}_c, \mathbf{E} = [\mathbf{E}^1, \dots, \mathbf{E}^m]^T, \\ & \quad \mathcal{Z} = \Phi(\mathbf{Z}^1, \dots, \mathbf{Z}^m), C = \Psi(\mathbf{C}^1, \dots, \mathbf{C}^m), \end{aligned} \quad (8)$$

where α and γ are the trade-off parameters. Ψ denotes the merging operation, which merges $\{\mathbf{C}^v\}_{v=1}^m$ into the complementary tensor $C \in \mathbb{R}^{n \times c \times m}$. $\mathbf{E} = [\mathbf{E}^1, \dots, \mathbf{E}^m]^T$ is obtained by horizontally concatenating along the row of error matrices $\{\mathbf{E}^v\}_{v=1}^m$, which captures the outliers and noise.

Then, rather than using only the consistent part for subsequent clustering task [17], we employ a concatenation-fusion strategy to combine both the learned consistent tensor \mathcal{Z} and complementary tensor C to a unified embedding

$$\mathbf{U} = \frac{1}{m} \sum_{v=1}^m [\mathcal{Z}^v, \mathbf{C}^v] \in \mathbb{R}^{n \times 2c}. \quad (9)$$

By doing so, we leverage the useful information provided by both the consistent and complementary parts to enhance the discriminability of our proposed CFMVC-ETR model.

At last, we impose spectral clustering on the recovered affinity matrix $\mathbf{S} = \mathbf{U} \Sigma^{-1} \mathbf{U}^T \in \mathbb{R}^{n \times n}$ to obtain the final clustering result [14], where Σ is the diagonal matrix with the entry $\Sigma(i, i) = \sum_{j=1}^n \mathbf{U}(j, i)$.

Remark 1. [The superiority of ETR] Inspired by the Geman function [10], we design a novel approximation function $f_{ETR}(x) = \frac{e^{\delta^2} x}{\delta + x}$, $0 < \delta \leq 1$. Basically, $f_{ETR}(0) = 0$ is satisfied, which is consistent with the true rank function. To verify the superiority of the proposed ETR, we conduct comparisons with two representative tensor rank approximation methods, TNN [39] and LTS_pN [11]. As shown in Fig. 1 (a), ETR provides a much better approximation to the true rank than TNN and LTS_pN, especially for near-zero singular values. When $x \rightarrow 0$, $f_{ETR}(x) \gg x$ and $f_{ETR}(x) \gg \log(1 + x^p)$, which implies

that ETR penalizes near-zero singular values more effectively than TNN and TLS_pN . Therefore, with the minimization of $\|\mathcal{Z}\|_{ETR}$, our model enjoys a noise-robust property, which eliminates the noise and ensures that \mathcal{Z} learns the high-order consistency across views.

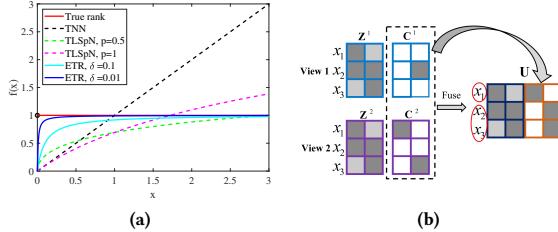


Figure 1: (a). Comparisons of different approximations methods. (b). An example for our CFMVC-ETR, $\{\mathbf{Z}^v\}_{v=1}^m$ and $\{\mathbf{C}^v\}_{v=1}^m$ are learned by Eq. (8), \mathbf{U} is obtained via Eq. (9).

Remark 2. [The benefit of TER and concatenation-fusion strategy] Unlike other methods such as HSIC [3] that employ matrix-level constraints to explore complementary information, our proposed TER is a tensor-level regularization that enables the exploration of high-order complementarity by treating all views as a whole. With γ increasing, the exclusivity of \mathbf{C}^v will be promoted, which enhances the view-specific information and captures the complementarity among views. As the example presented in Fig. 1(b), existing models that only use the learned consistent part $\{\mathbf{Z}^v\}_{v=1}^m$ for clustering task struggle to differentiate samples like x_2 that are on the boundary of similar classes. In contrast, our proposed concatenation-fusion strategy, which considers both the consistent part $\{\mathbf{Z}^v\}_{v=1}^m$ and complementary part $\{\mathbf{C}^v\}_{v=1}^m$, generates a more discriminative embedding \mathbf{U} that reveals a distinct clustering structure. The results on the Animal data set (as presented in Section 5) also verify the effectiveness of our concatenation-fusion strategy in partitioning similar classes.

4 OPTIMIZATION

Inspired by the alternation direction method of multipliers (ADMM) [18], we introduce the auxiliary tensor variable \mathcal{G} , so the model (8) can be rewritten as the following unconstrained problem,

$$\begin{aligned} \mathcal{L}(\{\mathbf{Z}^v\}_{v=1}^m, \{\mathbf{C}^v\}_{v=1}^m, \{\mathbf{H}^v\}_{v=1}^m, \mathcal{G}, \mathbf{E}, \{\mathbf{Y}^v\}_{v=1}^m, \mathcal{W}) \\ = \|\mathcal{G}\|_{ETR} + \alpha \|\mathbf{E}\|_{2,1} + \gamma \|\mathbf{C}\|_{TER} + \langle \mathcal{W}, \mathcal{Z} - \mathcal{G} \rangle \\ + \sum_{v=1}^m (\langle \mathbf{Y}^v, \mathbf{X}^v - (\mathbf{Z}^v + \mathbf{C}^v)\mathbf{H}^v - \mathbf{E}^v \rangle \\ + \frac{\mu}{2} \|\mathbf{X}^v - (\mathbf{Z}^v + \mathbf{C}^v)\mathbf{H}^v - \mathbf{E}^v\|_F^2) + \frac{\rho}{2} \|\mathcal{Z} - \mathcal{G}\|_F^2, \end{aligned} \quad (10)$$

where \mathcal{W} and $\{\mathbf{Y}^v\}_{v=1}^m$ are Lagrange multipliers, and μ and ρ are penalty parameters to control convergence. Then, we solve the variables in Eq. (10) through the following five subproblems.

4.1 $\{\mathbf{C}^v\}_{v=1}^m$ -Subproblem

With other variables fixed, the problem with \mathbf{C}^v is formulated as

$$\begin{aligned} \arg \min_{\{\mathbf{C}^v\}} \gamma \|\mathbf{C}\|_{TER} + \langle \mathbf{Y}^v, \mathbf{X}^v - (\mathbf{Z}^v + \mathbf{C}^v)\mathbf{H}^v - \mathbf{E}^v \rangle \\ + \frac{\mu}{2} \|\mathbf{X}^v - (\mathbf{Z}^v + \mathbf{C}^v)\mathbf{H}^v - \mathbf{E}^v\|_F^2. \end{aligned} \quad (11)$$

The derivative of Eq. (11) is set to 0, we have

$$\begin{bmatrix} \text{vec}(\mathbf{C}^1) \\ \vdots \\ \text{vec}(\mathbf{C}^m) \end{bmatrix} = \mathbf{P}^\dagger \begin{bmatrix} \text{vec}(\mathbf{B}^1) \\ \vdots \\ \text{vec}(\mathbf{B}^m) \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \mu & \gamma & \cdots & \gamma \\ \gamma & \mu & \cdots & \gamma \\ \vdots & \vdots & \ddots & \vdots \\ \gamma & \gamma & \cdots & \mu \end{bmatrix}. \quad (12)$$

where $\text{vec}(\cdot)$ denotes matrix vectorization. $\mathbf{B}^v = \mathbf{Y}^v \mathbf{H}^{vT} + \mu \mathbf{X}^v \mathbf{H}^{vT} - \mu \mathbf{Z}^v - \mu \mathbf{E}^v \mathbf{H}^{vT}$, and \mathbf{P}^\dagger is the generalized inverse matrix of \mathbf{P} .

4.2 $\{\mathbf{Z}^v\}_{v=1}^m$ -Subproblem

Fixing the other variables leads to the following problem for \mathbf{Z}^v ,

$$\mathbf{Z}^{v*} = \arg \max_{\mathbf{Z}^v, \mathbf{Z}^v \mathbf{T} \mathbf{Z}^v = \mathbf{I}_c} \text{Tr}(\mathbf{Z}^{vT} \mathbf{M}^v), \quad (13)$$

where $\mathbf{M}^v = (\mu \mathbf{X}^v + \mathbf{Y}^v - \mu \mathbf{E}^v) \mathbf{H}^{vT} - \mu \mathbf{C}^v + \rho \mathbf{G}^v - \mathbf{W}^v$. The optimal solution of \mathbf{Z}^v is $\mathbf{U}_Z^v \mathbf{V}_Z^{vT}$, where \mathbf{U}_Z^v and \mathbf{V}_Z^v are the left and right singular matrix of \mathbf{M}^v .

4.3 E-Subproblem

Fixing the other variables, the problem with \mathbf{E} is formulated as,

$$\arg \min_{\mathbf{E}} \frac{\alpha}{\mu} \|\mathbf{E}\|_{2,1} + \frac{1}{2} \|\mathbf{E} - \hat{\mathbf{E}}\|_F^2, \quad (14)$$

where $\hat{\mathbf{E}}^T$ is constructed by horizontally concatenating the matrices $\mathbf{X}^v - (\mathbf{Z}^v + \mathbf{C}^v)\mathbf{H}^v - \mathbf{E}^v + \frac{1}{\mu} \mathbf{Y}^v$ together along row. Its solution can be obtained by $\ell_{2,1}$ minimization thresholding operator as in [21],

$$\mathbf{E}_{i,:} = \begin{cases} \frac{\|\hat{\mathbf{E}}_{i,:}\|_2 - \frac{\alpha}{\mu}}{\|\hat{\mathbf{E}}_{i,:}\|_2} \hat{\mathbf{E}}_{i,:}, & \|\hat{\mathbf{E}}_{i,:}\|_2 > \frac{\alpha}{\mu}, \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

where $\hat{\mathbf{E}}_{i,:}$ is the i -th row of $\hat{\mathbf{E}}$.

4.4 G-Subproblem

When other variables are fixed, the subproblem for \mathcal{G} is formulated as,

$$\arg \min_{\mathcal{G}} \frac{1}{\rho} \|\mathcal{G}\|_{ETR} + \frac{1}{2} \left\| \mathcal{G} - \left(\mathcal{Z} + \frac{\mathcal{G}}{\rho} \right) \right\|_F^2. \quad (16)$$

We refer to this problem as the Enhanced Tensorial Rank Minimization problem (ETRM), which can be solved by the following theorem.

Theorem 1. Suppose $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with t-SVD $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ and $\beta > 0$, the Enhanced Tensorial Rank Minimization problem (ETRM) can be described as follows,

$$\arg \min_{\mathcal{G}} \beta \|\mathcal{G}\|_{ETR} + \frac{1}{2} \|\mathcal{G} - \mathcal{A}\|_F^2. \quad (17)$$

Then, optimal solution \mathcal{G}^* is obtained as,

$$\mathcal{G}^* = \mathcal{U} * \text{ifft}(\text{Proxf}, \beta(\mathcal{S}_f), [], 3) * \mathcal{V}^T, \quad (18)$$

where $\text{ifft}(\text{Proxf}, \beta(\mathcal{S}_f), [], 3) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f -diagonal tensor, and $\text{Proxf}, \beta(\mathcal{S}_f^k(i, i))$ satisfies the following equation,

$$\text{Proxf}, \beta(\mathcal{S}_f^k(i, i)) = \arg \min_{x \geq 0} \frac{1}{2} (x - \mathcal{S}_f^k(i, i))^2 + \beta f(x), \quad (19)$$

where $f(x) = \frac{e^{\delta^2 x}}{\delta + x}$.

The proof of Theorem 1 is given in the supplementary. Eq. (19) is a combination of concave and convex functions, so we can use the difference of convex (DC) programming [29] to acquire a closed-form solution,

$$\tau^{iter+1} = \left(S_f^k(i, i) - \frac{\partial f(\tau^{iter})}{\rho} \right)_+, \quad (20)$$

where $\tau = \text{Prox}_{f,\beta}(S_f^k(i, i))$, $f(x) = \frac{e^{\delta^2 x}}{x+\gamma}$ and $iter$ is the number of iterations.

4.5 $\{\mathbf{H}^v\}_{v=1}^m$ -Subproblem

Fixing the other variables, \mathbf{H}^v can be updated by,

$$\mathbf{H}^{v*} = \arg \max_{\mathbf{H}^v \mathbf{H}^{vT} = \mathbf{I}_c} \text{Tr}(\mathbf{H}^{vT} \mathbf{N}^v), \quad (21)$$

where $\mathbf{N}^v = (\mathbf{Z}^v + \mathbf{C}^v)^T (\mu \mathbf{X}^v + \mathbf{Y}^v - \mu \mathbf{E}^v)$. The optimal solution of \mathbf{H}^v is $\mathbf{U}_H^v \mathbf{V}_H^v$, where \mathbf{U}_H^v and \mathbf{V}_H^v are the left and right singular matrix of \mathbf{N}^v .

At last, the Lagrange multipliers and penalty parameters are updated as follows,

$$\begin{cases} \mathbf{Y}^v = \mathbf{Y}^v + \mu(\mathbf{X}^v - (\mathbf{Z}^v + \mathbf{C}^v)\mathbf{H}^v - \mathbf{E}^v) \\ \mathcal{W} = \mathcal{W} + \rho(\mathcal{Z} - \mathcal{G}) \\ \mu = \eta_\mu \mu, \mu = \min(\mu, \mu_{\max}) \\ \rho = \eta_\rho \rho, \rho = \min(\eta, \rho_{\max}) \end{cases} \quad (22)$$

where $\eta_\mu, \eta_\rho > 1$ are used to accelerate convergence. The complete procedure is summarized in Algorithm 1.

Algorithm 1 Optimization Algorithm of CFMVC-ETR

Input: Multi-view data $\{\mathbf{X}^1, \dots, \mathbf{X}^m\}$, cluster number c , trade-off parameters γ and α .

Initialize: Set $\mathbf{H}^v, \mathbf{Z}^v, \mathbf{E}^v, \mathbf{Y}^v$ to zero matrix, $\mathcal{G} = \mathcal{W} = \mathbf{0}$, $\mu = 10^{-5}, \rho = 10^{-4}, \eta_\mu = \eta_\rho = 2, \mu_{\max} = \rho_{\max} = 10^{10}, \epsilon = 10^{-7}$.

- 1: **while** not converge **do**
- 2: Update \mathbf{C}^v by Eq. (12);
- 3: Update \mathbf{Z}^v by Eq. (13);
- 4: Update \mathbf{E} by Eq. (15);
- 5: Update \mathcal{G} by Eq. (20);
- 6: Update \mathbf{H}^v by Eq. (21);
- 7: Update μ, ρ, \mathbf{Y}^v , and \mathcal{W} by Eq.(22);
- 8: Check the convergence conditions:
 $\|\mathbf{X}^v - (\mathbf{Z}^v + \mathbf{C}^v)\mathbf{H}^v - \mathbf{E}^v\|_\infty < \epsilon \& \|\mathcal{Z} - \mathcal{G}\|_\infty < \epsilon$
- 9: **end while**
- 10: Perform spectral clustering on \mathbf{S} to obtain the final clustering results.

4.6 Convergence Analysis

The ADMM algorithms are usually employed to solve convex problems with two variables. When it comes to dealing with non-convex and multi-block problems, convergence is hard to justify. However, for our CFMVC-ETR, the convergence of Algorithm 1 is theoretically guaranteed by the Theorem 2 and the proof has been given in the supplementary.

Theorem 2. Let $\{\mathcal{P}_k = (\mathbf{Z}_k^v, \mathbf{C}_k^v, \mathbf{E}_k^v, \mathbf{H}_k^v, \mathbf{Y}_k^v, \mathcal{W}_k, \mathcal{G}_k)\}_{k=1}^\infty$ be the sequence generated by Algorithm 1, then the sequence $\{\mathcal{P}_k\}_{k=1}^\infty$ satisfies the following two principles:

- 1). $\{\mathcal{P}_k\}_{k=1}^\infty$ is bounded.
- 2). Any accumulation point of $\{\mathcal{P}_k\}_{k=1}^\infty$ is a KKT point of Eq. (10).

4.7 Complexity Analysis

Time complexity: The mean computational complexity of our CFMVC-ETR focuses on two parts. The first part is to update the variables (i.e., $\mathbf{C}^v, \mathbf{Z}, \mathbf{E}^v, \mathcal{G}, \mathbf{H}^v$) in problem (10), which take $\mathcal{O}(ncd^v + m^3 + m^2nc), \mathcal{O}(d^{v3} + nd^{v2} + ncd^v), \mathcal{O}(nd), \mathcal{O}(mnc \log(mn) + m^2nc)$ and $\mathcal{O}(d^{v3} + nd^{v2})$, where $d = \sum_{v=1}^m d^v$, respectively. The second part is performing spectral clustering on \mathbf{S} , which costs $\mathcal{O}(m^3c^3 + m^2c^2n)$ by using the method in [14]. For $c \ll n$ and $m \ll n$, the time complexity of CFMVC-ETR is $\mathcal{O}(mcn \log(mn) + n(d_{\max})^2), d_{\max} = \max_{v=1}^m (d^v)$, which is considerably lower than the $\mathcal{O}(n^2 \log(n))$ of the existing tensor method.

Space complexity: The major memory costs of CFMVC-ETR are matrices $\mathbf{C}^v, \mathbf{Z}^v, \mathbf{E}^v, \mathbf{H}^v$ and tensor \mathcal{G} and \mathcal{W} . Consequently, the space complexity of our proposed method is $\mathcal{O}((d + mc)n)$, which is linear to the sample size.

Table 1: Details of the used data sets.

Data set	Type	Sample	Cluster	View
BBCSport	Text	544	5	2
NGs	Text	500	5	3
Caltech101-all	Object	9144	102	6
Alois-100	Object	11025	100	4
Animal	Object	11673	20	4
CIFAR10	Object	50000	10	3
Noisy MNIST	Digit	50000	10	2
CCV	Video	6773	20	3
YoutubeFace	Video	101499	31	5

5 EXPERIMENT

5.1 Experimental Settings

Data sets: Nine challenging data sets are adopted to validate our CFMVC-ETR, including text data sets NGs and BBCSport, object data sets Caltech101-all, Alois-100, Animal, and CIFAR10, digit data set Noisy MNIST, video data sets CCV and YoutubeFace. More details can be found in Table 1.

Baselines: To validate the superiority of CFMVC-ETR, nine state-of-the-art multi-view clustering methods are selected for comparison, including MVGL(2017) [42], MVC-DMF-PA (2021) [44], EOMSC-CA (2022) [20], SFMC (2020) [16], SMVSC (2021) [25], t-SVD-MSC (2018) [39], ETLMSC (2019) [36], and TBGL (2022) [38], respectively. Furthermore, we also perform standard spectral clustering (SC) [31] on each view and show the best results.

Evaluation Metrics: We employ four commonly used metrics to measure the clustering quality, including accuracy (ACC), normalized mutual information (NMI), purity (PUR), and adjusted rand index (ARI). For all metrics, the larger values indicate better performances.

Parameter Setting: For our CFMVC-ETR, there are three parameters that need to be tuned, α and γ are trade-off parameters

Table 2: Results (mean(std)) of our proposed method and other compared methods on nine data sets. 'OM' indicates the “out-of-memory error”.

Data set	Metric (%)	<i>SC_{best}</i>	MVGL	MVC-DMF-PA	EOMSC-CA	SFMC	SMVSC	t-SVD-MSC	ETLMS	TBGL	Ours
NGs	ACC	25.90(0.25)	22.80(0)	44.20(1.58)	63.00(0)	22.20(0)	74.20(0)	90.00(0)	24.46(1.38)	32.20(0)	99.40(0)
	NMI	1.81(0.13)	6.87(0)	27.31(1.47)	47.85(0)	2.87(0)	52.38(0)	<u>76.64(0)</u>	5.80(1.66)	9.35(0)	98.08(0)
	PUR	25.90(0.25)	23.80(0)	46.40(1.56)	64.80(0)	22.80(0)	74.20(0)	<u>81.84(0)</u>	20.06(0.25)	32.80(0)	99.40(0)
	ARI	0.52(0.11)	0.21(0)	13.24(1.06)	41.30(0)	0.18(0)	47.25(0)	<u>77.57(0)</u>	0.50(0.57)	9.32(0)	98.50(0)
BBCSport	ACC	50.35(0.35)	22.8(0)	<u>59.74(3.69)</u>	46.32(0)	36.40(0)	59.01(0)	34.38(0)	50.17(2.64)	54.78(0)	99.26(0)
	NMI	21.00(0.22)	6.87(0)	<u>31.18(2.28)</u>	21.73(0)	1.75(0)	30.07(0)	1.45(0)	<u>31.97(0.98)</u>	27.75(0)	97.40(0)
	PUR	55.31(0.35)	23.80(0)	61.03(2.48)	50.37(0)	36.95(0)	<u>61.40(0)</u>	23.60(0)	40.65(0.71)	54.96(0)	99.26(0)
	ARI	15.93(0.45)	0.21(0)	21.70(2.43)	15.04(0)	0.56(0)	27.63(0)	-0.65(0)	22.64(1.71)	18.75(0)	98.41(0)
CCV	ACC	13.69(0.22)	11.32(0)	16.74(0.48)	24.32(0)	15.49(0)	22.52(0)	47.80(0.04)	19.23(0.49)	15.13(0)	74.96(0)
	NMI	9.51(0.12)	3.15(0)	12.29(0.30)	18.70(0)	8.14(0)	16.18(0)	<u>41.34(0.05)</u>	15.42(0.34)	7.42(0)	76.97(0)
	PUR	17.50(0.15)	11.68(0)	20.51(0.36)	26.68(0)	17.05(0)	25.41(0)	<u>35.25(0.05)</u>	12.02(0.16)	16.34(0)	78.59(0)
	ARI	3.25(0.13)	0.06(0)	4.05(0.20)	8.52(0)	3.17(0)	7.17(0)	29.61(0.05)	6.30(0.14)	1.45(0)	61.81(0)
Caltech101-all	ACC	19.50(0.86)	14.34(0)	10.82(1.70)	24.70(0)	23.69(0)	29.18(0)	<u>47.45(1.01)</u>	21.65(1.11)	23.20(0)	57.63(0)
	NMI	40.57(0.33)	25.14(0)	20.28(5.53)	27.09(0)	23.50(0)	40.39(0)	<u>71.49(0.44)</u>	43.93(0.27)	21.39(0)	81.95(0)
	PUR	40.61(0.52)	22.86(0)	17.02(1.97)	27.64(0)	31.86(0)	38.17(0)	<u>53.74(1.88)</u>	36.71(2.76)	29.45(0)	81.47(0)
	ARI	13.44(0.92)	-0.71(0)	2.35(4.99)	10.02(0)	-0.32(0)	23.27(0)	<u>31.60(1.06)</u>	19.21(2.01)	0.16(0)	41.45(0)
Aloï-100	ACC	64.54(1.15)	53.44(0)	15.89(0.51)	23.70(0)	68.00(0)	34.35(0)	<u>71.99(1.44)</u>	66.82(2.25)	68.58(0)	78.93(0)
	NMI	80.03(0.33)	67.22(0)	30.66(0.60)	57.93(0)	70.14(0)	60.99(0)	<u>83.92(0.43)</u>	81.67(0.53)	72.20(0)	91.15(0)
	PUR	67.28(0.94)	55.94(0)	17.40(0.57)	24.92(0)	69.06(0)	35.58(0)	58.03(1.56)	53.05(2.73)	69.81(0)	81.69(0)
	ARI	54.29(1.08)	4.55(0)	4.89(0.40)	7.42(0)	11.36(0)	20.39(0)	61.71(1.13)	57.78(2.06)	14.49(0)	74.86(0)
Animal	ACC	14.35(0.19)	9.34(0)	12.88(0.21)	18.87(0)	9.17(0)	19.49(0)	17.16(1.01)	12.95(0.39)	9.16(0)	93.78(0)
	NMI	10.47(0.12)	0.84(0)	9.36(0.18)	14.80(0)	0.17(0)	<u>15.72(0)</u>	12.81(0.62)	8.38(0.70)	0.17(0)	93.06(0)
	PUR	17.84(0.20)	9.59(0)	17.01(0.34)	21.37(0)	9.22(0)	<u>21.34(0)</u>	10.56(0.13)	6.71(0.17)	9.21(0)	95.73(0)
	ARI	3.24(0.10)	0.01(0)	2.73(0.16)	6.99(0)	0.00(0)	<u>8.35(0)</u>	4.73(0.02)	1.73(0.32)	0(0)	92.11(0)
CIFAR10	ACC	89.41(4.06)	OM	<u>99.21(8.74)</u>	99.08(0)	97.77(0)	98.35(0)	OM	OM	OM	100(0)
	NMI	80.08(1.41)	OM	<u>97.84(4.06)</u>	97.48(0)	94.63(0)	97.03(0)	OM	OM	OM	100(0)
	PUR	89.67(3.25)	OM	<u>99.21(6.32)</u>	99.08(0)	97.77(0)	98.85(0)	OM	OM	OM	100(0)
	ARI	79.79(3.00)	OM	<u>98.27(8.11)</u>	97.97(0)	95.14(0)	97.50(0)	OM	OM	OM	100(0)
Noisy MNIST	ACC	<u>61.10(2.78)</u>	OM	56.25(1.56)	58.65(0)	41.12(0)	18.44(0)	OM	OM	OM	99.49(0)
	NMI	<u>61.82(1.58)</u>	OM	48.71(0.71)	49.08(0)	34.81(0)	11.77(0)	OM	OM	OM	98.56(0)
	PUR	<u>68.34(1.23)</u>	OM	58.35(1.19)	58.83(0)	45.35(0)	22.97(0)	OM	OM	OM	99.49(0)
	ARI	<u>50.02(2.52)</u>	OM	38.79(0.86)	40.44(0)	16.55(0)	6.02(0)	OM	OM	OM	98.89(0)
YoutubeFace	ACC	OM	OM	OM	<u>26.66(0)</u>	OM	25.87(0)	OM	OM	OM	28.51(0)
	NMI	OM	OM	OM	<u>0.12(0)</u>	OM	<u>22.92(0)</u>	OM	OM	OM	24.95(0)
	PUR	OM	OM	OM	26.68(0)	OM	<u>33.21(0)</u>	OM	OM	OM	37.91(0)
	ARI	OM	OM	OM	0.01(0)	OM	<u>6.74(0)</u>	OM	OM	OM	9.73(0)

Table 3: Average running time (sec.) comparison on the data sets with more than 10,000 samples. 'OM' indicates the “out-of-memory error”.

Data set	MVGL	MVC-DMF-PA	EOMSC-CA	SFMC	SMVSC	t-SVD-MSC	ETLMS	TBGL	Ours
Caltech102-all	6006.82	361.69	<u>98.82</u>	7828.40	980.72	13841.21	15768.25	10401.31	95.89
Aloï-101	2808.92	9.76	86.39	160.21	658.18	14494.12	12723.68	11000.23	<u>42.35</u>
Animal	6839.55	510.10	29.63	38.29	136.19	16333.47	21506.21	39614.04	101.20
CIFAR10	OM	834.28	<u>97.29</u>	72.18	547.36	OM	OM	OM	193.38
Noisy MNIST	OM	504.36	99.57	74.30	528.41	OM	OM	OM	<u>84.56</u>
YoutubeFace	OM	OM	<u>570.94</u>	OM	1887.04	OM	OM	OM	246.34

to balance the importance of different terms in objective function, which both take the range of $\{10^{-4}, 10^{-2}, \dots, 10^3\}$. The parameter δ in ETR has a search range of $\{0.0001, \dots, 1\}$. For the baselines, we follow the parameter settings in the corresponding literature and report the best results. All the methods run 10 times and their averages and standard deviations are reported. All the experiments are implemented on a computer with a 2.50GHz i7-11700 CPU and 64GB RAM, Matlab R2021a.

5.2 Experimental Results

Table 2 displays the clustering performance of all methods on nine challenging data sets. The best and second best results in the table are denoted by **bold** value and underline value. From Table 2, we can conclude the following interesting observations.

1) From a global perspective, CFMVC-ETR outperforms all the baselines over four metrics on nine data sets. More importantly, CFMVC-ETR achieves the **desired ideal clustering performance on the CIFAR10 data set**. In addition, the improvement of clustering performance on some data sets is also remarkable. For example,

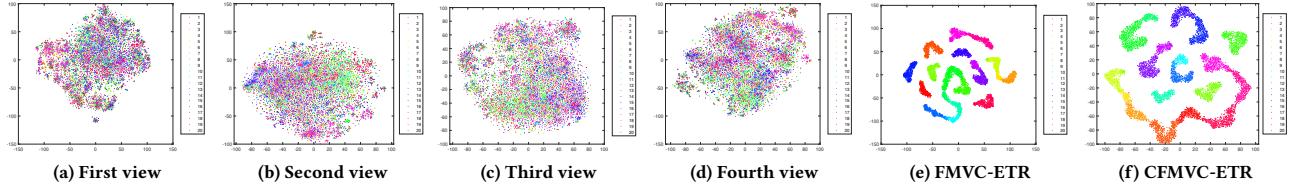


Figure 2: The t-SNE visualizations of the original features and the learned embedding of FMVC-ETR and CFMVC-ETR on the Animal data set.

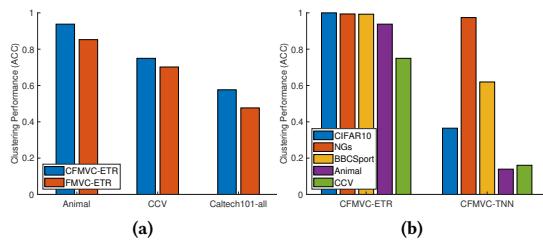


Figure 3: (a). The clustering performance (ACC) of FMVC-ETR and CFMVC-TNN on three data sets. (b). The clustering performance (ACC) of CFMVC-ETR and CFMVC-TNN on five data sets.

on the Animal data set, CFMVC-ETR outperforms the second-best SMVSC methods by **74.29%, 77.34%, 74.39%, and 83.76%**, respectively. For the Noisy MNIST data set, CFMVC-ETR gains improvements around **38.39%, 36.74%, 31.15%, and 48.87%** in terms of four metrics, respectively. These results provide strong evidence of the superiority of our algorithm. Meanwhile, CFMVC-ETR achieves satisfactory clustering performance for various data domains (e.g., text data, object data, video data, etc.), which further shows the clustering robustness under different scenarios.

2) Our proposed CFMVC-ETR and other tensor-based methods, i.e., t-SVD-MSC, ETLMSC, and TBGL outperform the non-tensor approaches in most cases. This is mainly due to the fact that tensor-based methods can well capture the high-order correlation, while non-tensor methods cannot.

3) Our proposed CFMVC-ETR is remarkably superior to all the other tensor-based methods on all nine data sets. We attribute such success to the fact that CFMVC-ETR simultaneously takes both consistency and complementarity into consideration, which makes the final embedding features more discriminative. In contrast, other tensor-based methods merely focus on consistency learning or only use the consistent part for the clustering task.

4) The tensor-based methods and several non-tensor methods such as MVGL and SFMC are not computationally feasible on all the large-scale data sets due to the out-of-memory error, especially on the YoutubeFace data set, which has over 100,000 samples. In contrast, our proposed CFMVC-ETR can effectively handle all the large-scale data sets (i.e., CIFAR10, Noisy MNIST, and YoutubeFace) and achieve satisfactory clustering performances, which benefits from the fact that our model enjoys a linear space complexity.

In order to further demonstrate the efficiency of CFMVC-ETR, we present the running time on large-scale data sets with more than 10,000 samples in Table 3. It is admitted that CFMVC-ETR takes more time than anchor-based methods such as EOMSC-CA, SFMC, and SMVSC. Nevertheless, the proposed CFMVC-ETR achieves

remarkable clustering results on all the large-scale data sets in the acceptable time range. Especially, CFMVC-ETR outperforms the existing tensor-based methods in terms of efficiency, which benefits from the matrix factorization technique that reduces the dimensions of the representation tensor.

5.3 Ablation Studies

To validate the influence of different components in our CFMVC-ETR, we conduct the following ablation studies:

Effectiveness of TER and concatenation-fusion strategy:

To demonstrate the effectiveness of the concatenation-fusion strategy and TER, we develop a variant model named FMVC-ETR by excluding the learning and integration steps of the complementary part. Fig. 3(a) shows the clustering results of FMVC-ETR and CFMVC-ETR on three challenging data sets. It is observed that the clustering performance is boosted with the help of the complementary part. Moreover, we use the t-SNE [30] to visualize the original features, and the learned embedding of FMVC-ETR and CFMVC-ETR on the Animal data set, respectively. As shown in Fig. 2, the samples in each view are mixed and difficult to be correctly classified. Although FMVC-ETR manages to segregate most classes, it still struggles to classify samples of similar classes. In contrast, our proposed CFMVC-ETR achieves a distinctive clustering structure, especially for the boundaries of similar classes. The above experimental results provide strong evidence for the effectiveness of our TER and concatenation-fusion strategy in capturing complementary information and improving clustering performance among similar classes.

Comparison with TNN: To validate the superiority of our proposed Enhanced Tensor Rank (ETR), we replace the ETR in the model (8) with the commonly used Tensor Nuclear Norm (TNN), deriving a variant of the proposed method denoted as CFMVC-TNN. Then, the clustering performance (ACC) on five challenging data sets is presented in Fig. 3(b). As illustrated, CFMVC-ETR outperforms CFMVC-TNN on all data sets. This can be attributed to the fact that TNN treats all singular values equally, resulting in the over-punishment of important components and the under-punishment of noises. In contrast, ETR assigns appropriate penalties to different singular values, enabling the removal of noise and preserving significant information.

5.4 Model Analysis

Influence of δ in ETR: Taking NGs, BBCSport, CCV, and Caltech101-all data sets as examples, we analyze the impact of δ in ETR. Specifically, we fix other parameters and report the clustering performance with different δ in the range of $\{10^{-4}, \dots, 1\}$ in Fig. 4. Notably, we

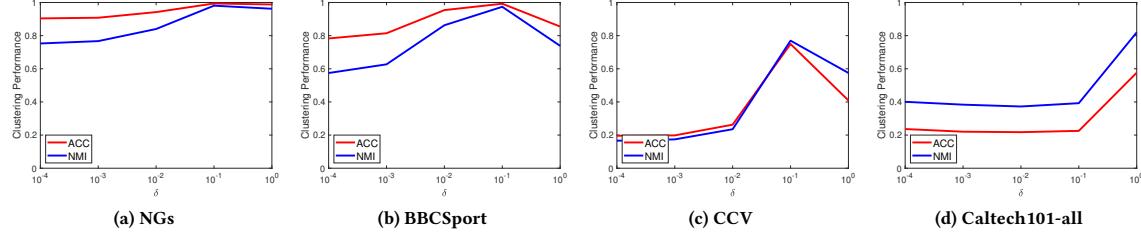
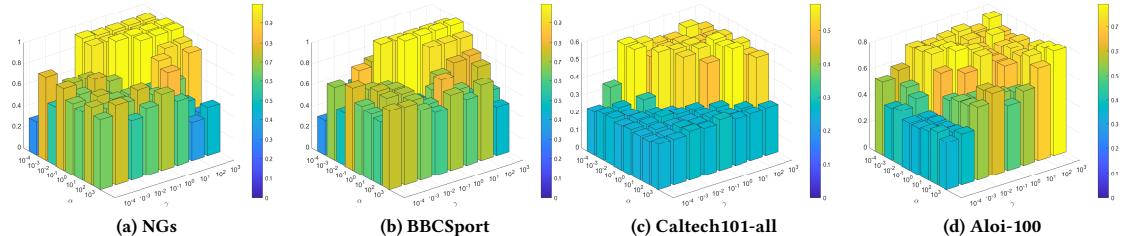
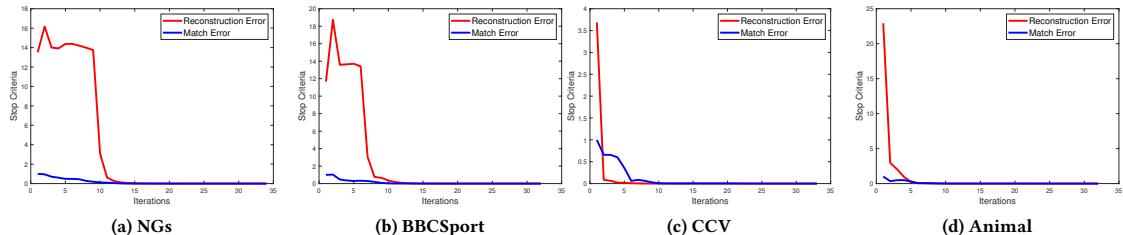
Figure 4: The performance (i.e., ACC and NMI) of CFMVC-ETR with varying parameter δ on four data sets.Figure 5: Parameters analysis: The clustering performances (ACC) with different parameters α and γ on four data sets.

Figure 6: Convergence Analysis: The stop criteria (i.e., RE and ME) variation curves on four data sets.

observe that the performance varies greatly for different δ . When $\delta = 0.1$, CFMVC-ETR achieves the best performance on NGs, BBC-Sport, and CCV data sets, but it peaks at $\delta = 1$ for the Caltech101-all data set. This phenomenon shows that δ plays a crucial role in our proposed ETR and exhibits variability across different data sets. This is because δ controls the penalty strength for different singular values, and different data sets usually contain different amounts of noise, leading to diverse demands on the strength of their penalties for small singular values.

Parameter Sensitivity Analysis: Here, we show the sensitivity of our proposed CFMVC-ETR to different trade-off parameters. CFMVC-ETR has two trade-off parameters α and γ that need to be tuned. Specifically, we vary α and γ over the range of $\{10^{-4}, \dots, 10^3\}$ and record the variations of the ACC metrics. The results for NGs, BBCSport, Caltech101-all, and Aloi-100 data sets are presented in Fig. 5. As shown, the parameters α and γ have a significant impact on clustering performance. And CFMVC-ETR can achieve favorable and stable clustering performance within a narrow range of parameters. Therefore, the search ranges of α and γ are limited to $\{10^{-4}, 10^{-3}, \dots, 10^0\}$ and $\{10^{-1}, 10^0, \dots, 10^3\}$, respectively.

Convergence Analysis: To further validate the convergence of our proposed CFMVC-ETR, we display the variations of the stop criteria on four data sets, and the stop criteria are denoted as Reconstruction Error (RE): $RE = \max_v \|\mathbf{X}^v - \mathbf{X}^v \mathbf{Z}^v - \mathbf{E}^v\|_\infty$ and Match Error (ME): $ME = \|\mathcal{Z} - \mathcal{G}\|_\infty$. As shown in Fig. 6, the values

of RE and ME rapidly converge to 0 within 15 steps and remain stable, indicating the promising convergence property. Thus, the convergence of Algorithm 1 can be proved not only theoretically (see Theorem 2), but also experimentally.

6 CONCLUSION

In this paper, we propose a fast tensorial multi-view clustering framework termed CFMVC-ETR, which integrates matrix factorization, tensor-level constraints, complementarity learning, and consistency learning into a unified optimization model. In particular, we use the matrix factorization technique to obtain two representation tensors, namely, the consistent tensor and the complementary tensor, and then design Enhanced Tensor Rank and Tensorial Exclusive Regularization for these two parts to explore high-order consistency and complementarity, respectively. Furthermore, we propose a concatenation-fusion strategy to integrate the consistent and complementary parts into a discriminative embedding for the clustering task. An efficient algorithm is proposed to solve the objective function, which enjoys both economical computational complexity and good convergence. The experimental results on nine challenging data sets demonstrate the superiority of our proposed model in multi-view clustering tasks.

ACKNOWLEDGMENTS

This work was supported by the Fundamental Research Funds for the Central Universities (No. 2022JBZY019).

REFERENCES

- [1] Michael W Berry, Murray Browne, Amy N Langville, V Paul Pauca, and Robert J Plemmons. 2007. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis* 52, 1 (2007), 155–173.
- [2] Thomas Blumensath. 2014. Sparse matrix decompositions for clustering. In *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 1163–1167.
- [3] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. 2015. Diversity-induced multi-view subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 586–594.
- [4] Man-Sheng Chen, Chang-Dong Wang, and Jian-Huang Lai. 2023. Low-Rank tensor based proximity learning for multi-View clustering. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2023), 5076–5090.
- [5] Yongyong Chen, Xiaolin Xiao, Chong Peng, Guangming Lu, and Yicong Zhou. 2021. Low-rank tensor graph learning for multi-view subspace clustering. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 1 (2021), 92–104.
- [6] James W Cooley and John W Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.* 19, 90 (1965), 297–301.
- [7] Chris Ding, Xiaofeng He, and Horst D Simon. 2005. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proceedings of the 2005 SIAM International Conference on Data Mining*. SIAM, 606–610.
- [8] Chris Ding, Tat Li, Wei Peng, and Haesun Park. 2006. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 126–135.
- [9] Quanxue Gao, Wei Xia, Zhizhen Wan, Deyan Xie, and Pu Zhang. 2020. Tensor-SVD based graph learning for multi-view subspace clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 3930–3937.
- [10] Donald Geman and Chengda Yang. 1995. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing* 4, 7 (1995), 932–946.
- [11] Jipeng Guo, Yanfeng Sun, Junbin Gao, Yongli Hu, and Baocai Yin. 2023. Logarithmic schatten- p norm minimization for tensorial multi-View subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 3 (2023), 3396–3410.
- [12] Shudong Huang, Yixi Liu, Yazhou Ren, Ivor W Tsang, Zenglin Xu, and Juncheng Lv. 2022. Learning smooth representation for multi-view subspace clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*. 3421–3429.
- [13] Shudong Huang, Yixi Liu, Ivor W Tsang, Zenglin Xu, and Juncheng Lv. 2022. Multi-View subspace clustering by joint measuring of consistency and diversity. *IEEE Transactions on Knowledge and Data Engineering* (2022), 1–12.
- [14] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. 2020. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 4412–4419.
- [15] Lusi Li and Haibo He. 2020. Bipartite graph based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering* 34, 7 (2020), 3111–3125.
- [16] Xuelong Li, Han Zhang, Rong Wang, and Feiping Nie. 2020. Multiview clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 1 (2020), 330–344.
- [17] Youwei Liang, Dong Huang, Chang-Dong Wang, and S Yu Philip. 2022. Multi-view graph learning by joint modeling of consistency and inconsistency. *IEEE Transactions on Neural Networks and Learning Systems* (2022), 1–15.
- [18] Zhouchen Lin, Risheng Liu, and Zhixun Su. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. *Advances in Neural Information Processing Systems* 24 (2011), 1–9.
- [19] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Li Liu, Siqi Wang, Weixuan Liang, and Jiayong Shi. 2021. One-pass multi-view clustering for large-scale data. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 12344–12353.
- [20] Suyuan Liu, Siwei Wang, Pei Zhang, Kai Xu, Xinwang Liu, Changwang Zhang, and Feng Gao. 2022. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 7576–7584.
- [21] Yongli Liu, Xiaojin Zhang, Guiying Tang, and Di Wang. 2019. Multi-view subspace clustering based on tensor Schatten- p norm. In *IEEE International Conference on Big Data (Big Data)*. IEEE, 5048–5055.
- [22] Canyi Lu, Jia Shi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. 2016. Tensor robust principal component analysis: exact recovery of corrupted low-rank tensors via convex optimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5249–5257.
- [23] Canyi Lu, Jia Shi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. 2019. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 4 (2019), 925–938.
- [24] Feiping Nie, Shenfei Pei, Rong Wang, and Xuelong Li. 2020. Fast clustering with co-clustering via discrete non-negative matrix factorization for image identification. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2073–2077.
- [25] Mengjing Sun, Pei Zhang, Siwei Wang, Sihang Zhou, Wenxuan Tu, Xinwang Liu, En Zhu, and Changjian Wang. 2021. Scalable multi-view subspace clustering with unified anchors. In *Proceedings of the 29th ACM International Conference on Multimedia*. 3528–3536.
- [26] Xiaoli Sun, Rui Zhu, Ming Yang, Xiujun Zhang, and Yuanyan Tang. 2022. Sliced sparse gradient induced multi-view subspace clustering via tensorial arctangent rank minimization. *IEEE Transactions on Knowledge and Data Engineering* (2022), 1–14.
- [27] Chang Tang, Zhenglai Li, Jun Wang, Xinwang Liu, Wei Zhang, and En Zhu. 2022. Unified one-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering* (2022), 1–11.
- [28] Yongqiang Tang, Yuan Xie, and Wensheng Zhang. 2023. Affine subspace robust low-rank self-representation from matrix to tensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023), 1–17.
- [29] Pham Dinh Tao and LT Hoai An. 1997. Convex analysis approach to DC programming: theory, algorithms and applications. *Acta Mathematica Vietnamica* 22, 1 (1997), 289–355.
- [30] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008), 2579–2605.
- [31] Ulrike Von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and Computing* 17, 4 (2007), 395–416.
- [32] Xinhang Wan, Jiyuan Liu, Weixuan Liang, Xinwang Liu, Yi Wen, and En Zhu. 2022. Continual Multi-view Clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*. 3676–3684.
- [33] Xinhang Wan, Xinwang Liu, Jiyuan Liu, Siwei Wang, Yi Wen, Weixuan Liang, En Zhu, Zhe Liu, and Lu Zhou. 2023. Auto-weighted multi-view clustering for large-scale data. *arXiv preprint arXiv:2303.01983* (2023), 1–9.
- [34] Jing Wang, Feng Tian, Hongchuan Yu, Chang Hong Liu, Kun Zhan, and Xiao Wang. 2017. Diverse non-negative matrix factorization for multiview data representation. *IEEE Transactions on Cybernetics* 48, 9 (2017), 2620–2632.
- [35] Jie Wen, Zheng Zhang, Zhao Zhang, Lei Zhu, Lunke Fei, Bob Zhang, and Yong Xu. 2021. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 10273–10281.
- [36] Jianlong Wu, Zhouchen Lin, and Hongbin Zha. 2019. Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing* 28, 12 (2019), 5910–5922.
- [37] Yanan Wu, Tengfei Liang, Songhe Feng, Yi Jin, Gengyu Lyu, Haojun Fei, and Yang Wang. 2023. MetaZSCIL: A meta-learning approach for generalized zero-shot class incremental learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 10408–10416.
- [38] Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. 2023. Tensorized bipartite graph learning for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 4 (2023), 5187–5202.
- [39] Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu. 2018. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision* 126, 11 (2018), 1157–1179.
- [40] Weiqing Yan, Jindong Xu, Jinglei Liu, Guanghui Yue, and Chang Tang. 2022. Bipartite graph-based discriminative feature learning for multi-view clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*. 3403–3411.
- [41] Haizhou Yang, Quanxue Gao, Wei Xia, Ming Yang, and Xinbo Gao. 2022. Multiview spectral clustering with bipartite graph. *IEEE Transactions on Image Processing* 31 (2022), 3591–3605.
- [42] Kun Zhan, Changqing Zhang, Junpeng Guan, and Junsheng Wang. 2017. Graph learning for multi-view clustering. *IEEE Transactions on Cybernetics* 48, 10 (2017), 2887–2895.
- [43] Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao. 2015. Low-rank tensor constrained multiview subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1582–1590.
- [44] Chen Zhang, Siwei Wang, Jiyuan Liu, Sihang Zhou, Pei Zhang, Xinwang Liu, En Zhu, and Changwang Zhang. 2021. Multi-view clustering via deep matrix factorization and partition alignment. In *Proceedings of the 29th ACM International Conference on Multimedia*. 4156–4164.
- [45] Tiejian Zhang, Xinwang Liu, En Zhu, Sihang Zhou, and Zhibin Dong. 2022. Efficient anchor learning-based multi-view clustering—a late fusion method. In *Proceedings of the 30th ACM International Conference on Multimedia*. 3685–3693.
- [46] Yi Zhang, Xinwang Liu, Siwei Wang, Jiyuan Liu, Sisi Dai, and En Zhu. 2021. One-stage incomplete multi-view clustering via late fusion. In *Proceedings of the 29th ACM International Conference on Multimedia*. 2717–2725.
- [47] Handong Zhao, Zhengming Ding, and Yun Fu. 2017. Multi-view clustering via deep matrix factorization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31. 2921–2927.
- [48] Pan Zhou, Canyi Lu, Jia Shi Feng, Zhouchen Lin, and Shuicheng Yan. 2019. Tensor low-rank representation for data recovery and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 5 (2019), 1718–1732.