Anchors Crash Tensor: Efficient and Scalable Tensorial Multi-view Subspace Clustering

Jintian Ji, Songhe Feng

Abstract—Tensorial Multi-view Clustering (TMC), a prominent approach in multi-view clustering, leverages low-rank tensor learning to capture high-order correlation among views for consistent clustering structure identification. Despite its promising performance, the TMC algorithms face three key challenges: 1). The severe computational burden makes it difficult for TMC methods to handle large-scale datasets. 2). Estimation bias problem caused by the convex surrogate of the tensor rank. 3). Lack of explicit balance of consistency and complementarity. Being aware of these, we propose a basic framework Efficient and Scalable Tensorial Multi-View Subspace Clustering (ESTMC) for large-scale multi-view clustering. ESTMC integrates anchor representation learning and non-convex function-based low-rank tensor learning with a Generalized Non-convex Tensor Rank (GNTR) into a unified objective function, which enhances the efficiency of the existing subspace-based TMC framework. Furthermore, a novel model ESTMC-C^2 with the proposed Enhanced Tensor Rank (ETR), Consistent Geometric Regularization (CGR), and Tensorial Exclusive Regularization (TER) is extended to balance the learning of consistency and complementarity among views, delivering divisible representations for the clustering task. Efficient iterative optimization algorithms are designed to solve the proposed ESTMC and ESTMC-C^2, which enjoy time-economical complexity and exhibit theoretical convergence. Extensive experimental results on various datasets demonstrate the superiority of the proposed algorithms as compared to state-of-the-art methods.

Index Terms—Multi-view clustering, subspace clustering, low-rank tensor learning, anchor representation learning

1 Introduction

Clustering is a basic task in unsupervised learning, which aims to partition the data into a certain number of groups by exploring the information hidden in the data. Many well-known algorithms such as k-means [1] and spectral clustering [2] perform well on single-view data. However, single-view data cannot provide a comprehensive depiction of objects, which limits the performance of these single-view clustering algorithms. To address this limitation, the exploration of multi-view data, which presents objects from diverse sources or perspectives, has garnered significant interest. Multi-view clustering [3]–[5] aims to exploit both consistent and complementary information across multiple views, thereby enhancing the understanding of the objects and improving clustering performance.

In recent years, many multi-view algorithms have emerged, which can be briefly divided into two categories, matrix-oriented multi-view clustering and Tensorial Multi-view Clustering (TMC). The matrix-oriented multi-view clustering methods [6]–[15] usually explore the information in individual views or between pairs of views. Depending on the method used to construct the representation matrices, these algorithms are usually classified as subspace-based methods, graph-based methods, and kernel-based methods, respectively. For subspace-based methods [16]–[21], they usually adopt the self-representation strategy to build representations that reveal the linear structure between samples. For example, Cao *et al.* [16] learns the low-dimensional subspace representation of each view and explores the respective smooth and diverse structures. The graph Laplacian regularization is employed in [18] to preserve the geometric features of the graph in each view. [19] imposes rank constraints on the Laplace matrix of each view to directly obtain the indicator matrices, and then uses spectral rotation to achieve the fusion of multiple views. To improve the efficiency of processing large-scale data, some works learn a set of anchors that cover all samples to reduce the size of representation matrices. For instance, Kang *et al.* [20] pre-select certain anchors to learn a smaller affinity matrix for clustering. [21], [22] integrate anchor selection and representation learning into a unified framework to avoid suboptimal solutions. Graph-based methods [7], [8], [23] explore the non-linear structure between samples by constructing similarity graphs. For example, [7] learns view-specific similarity graphs to construct a unified graph that Leads to direct response clustering results. To reduce the complexity of building graphs and clustering, [8], [23] introduce anchors to construct bipartite graphs in the graph learning framework. Kernel-based methods [24], [25] are another classic popular approach, which aims to explore heterogeneous information from multiple base kernels to improve clustering performance. For example, Liu *et al.* [25] propose a contrastive multi-view kernel generation method to integrate the views into a quality kernel to ensure their diversity and heterogeneity. Since these matrix-oriented algorithms split the multi-view problem into several single-view problems or one unified-view problem, it is difficult to explore the high-order correlation among all the views to obtain a consistent clustering structure.

• *Jintian Ji, Songhe Feng are with the School of Computer Science and Technology, Beijing Jiaotong University, Beijing, 100044, China, and Key Laboratory of Big Data & Artificial Intelligence in Transportation (Beijing Jiaotong University), Ministry of Education.*
E-mail: {22120385, shfeng}@bjtu.edu.cn.

**Corresponding Author: Songhe Feng*

Due to the challenges of exploring correlation among views with matrix-oriented constraints, recent methods like TMC approaches [11], [26]–[29] have emerged, focusing on capturing the high-order correlation among views. Specifically, these algorithms combine the representation matrices of different views into a third-order tensor and then leverage the low-rank tensor constraint for consistent clustering structure discovery. For example, [30] constructs a three-order tensor from multi-view matrices and employs a slice-based nuclear norm to exploit high-order correlation. To better constrain the low-rankness of the representation tensor, [31]–[33] introduce the tensor average rank for the representation tensor, with the Tensor Nuclear Norm (TNN) based on the tensor Singular Value Decomposition (t-SVD) serving as the tightest convex approximation. [26], [34] adopt the Tensor Nuclear Norm (TNN) on rotated representation tensor to improve the efficiency of capturing high-order correlation. However, subsequent studies have criticized TNN for its loose and biased approximation of tensor rank, as it over-penalizes large singular values and under-penalizes small ones. To address this problem, tensor ranks based on non-convex approximation functions have been proposed. For example, [11], [27] propose the weighted tensor nuclear norm that imposes different penalty parameters for different singular values. [35] proposes a Tensor Logarithmic Schatten-p Norm (TLS$_p$N), [36] constructs the Tensorial Arctangent Rank (TAR) based on the characteristics of the arctan-function. [37] proposes a generalized non-convex low-rank tensor approximation to integrate some common non-convex approximation functions. To mitigate the computational burden associated with tensor operations (*e.g.,* the tensor Singular Value Decomposition (t-SVD) and the Fast Fourier transformations (FFT)), [10], [38] pre-select a portion of the anchors to build bipartite graphs, which leads to a space-saving representation tensor.

Despite its progress, the TMC approach encounters several limitations: 1) **The tensor rank approximation problem.** Although non-convex tensor rank methods exhibit superior performance over TNN, an optimal formulation remains elusive. This unresolved issue calls for further research to identify a practical tensor rank model. 2) **High computational complexity problem.** The reliance on tensor operations often leads to substantial computational overhead in TMC methods. Moreover, pre-selected anchors introduce additional time and may compromise the optimality of solutions. 3) **The problem of balancing consistency and complementarity**. Many existing TMC algorithms inadequately address complementary information. Accurately clustering samples, especially those with similar classes or on class boundaries, can be challenging without explicitly capturing complementary aspects across views.

To address the aforementioned problems, we first propose a basic framework named Efficient and Scalable Tensorial Multi-View Subspace Clustering Framework (ESTMC) for large-scale TMC tasks (See Fig. 1), which integrates anchor representation learning and non-convex function-based low-rank tensor learning into a unified optimization model. Specifically, ESTMC adopts the anchor representation strategy to simultaneously learn optimal anchors and construct a space-saving representation tensor, a Generalized Non-convex Tensor Rank (GNTR) to explore the high-order correlation among views. In addition, we extend ESTMC with three tailored items (Enhanced Tensor rank (ETR), Consistent Geometric Regularization (CGR), and Tensorial Exclusive Regularization (TER)) to a novel model ESTMC-C^2 (See Fig. 3), which can explicitly capture the consistent parts and complementary parts in different views. Then a concatenation-fusion strategy is designed to integrate these two parts into discriminative representations. The novelty and main contributions of this paper are summarized as follows:

- ESTMC first integrates the anchor representation strategy with non-convex function-based low-rank tensor learning into a unified framework that can efficiently handle large-scale multi-view clustering problems. Furthermore, we give the standard form of the non-convex tensor rank (Generalized Non-convex tensor Rank (GNTR)), allowing ESTMC to accelerate most current TMC frameworks.
- ESTMC-C^2 proposes a novel Enhanced Tensor Rank (ETR) with an improved non-convex function to approximate the tensor rank, which leans to protect larger singular values and pushes smaller singular values to zero, thus a compact low-rank representation tensor can be desired.
- ESTMC-C^2 proposes Tensorial Exclusive Regularization (TER) and Consistent Geometric Regularization (CGR) to obtain the complementary information and consistent non-linear structure, which balances the consistency and complementarity of multiple views. Then a concatenation-fusion approach is designed to improve the discriminative properties of the sample representations.
- Efficient algorithms are proposed to solve ESTMC and ESTMC-C^2, which are time-economical and enjoy theoretically proven convergence. Experimental results over twelve datasets indicate that ESTMC and ESTMC-C^2 are superior to the state-of-the-art methods.

This paper is a significant extension of our conference work [39]. Compared with [39], several substantial improvements are concluded here: 1) We propose an efficient and scalable framework (ESTMC) to address the large-scale TMC problem. With an improved anchor representation strategy and the proposed GNTR, ESTMC is a versatile tool that can enhance the efficiency of most TMC approaches. 2) We propose a novel model ESTMC-C^2 to improve the discriminative properties of the sample representations. ESTMC-C^2 adopts the Enhanced Tensor Rank, Consistent Geometric Regularization (CGR), and Tensorial Exclusive Regularization (TER) to explicitly capture consistent and complementary information in different views, respectively. A concatenation-fusion strategy is designed to integrate the obtained consistent and complementary parts into discriminative representations for subsequent clustering tasks. 3) More comparison experiments are conducted on twelve challenging datasets to demonstrate the superiority of our proposed algorithms compared to competing methods.

The remainder of this paper is organized as follows. Section 2 presents a description of the notations and preliminaries. Section 3 details our proposed ESTMC and ESTMC-

C^2. In Section 4, we present the convergence analysis and complexity analysis of the proposed methods. The experimental results and analysis are shown in Section 5. Finally, Section 6 concludes this paper.

2 NOTATIONS AND RELATED WORKS

2.1 Notations

We summarize the frequently used notations in Table 1.

TABLE 1: Summary of notations.

Symbol	Definition
x, $\boldsymbol{x}$, $\mathbf{X}$, and $\mathcal{X}$	scalar, vector, matrix, and tensor
$\mathcal{X}^k$	the k-th frontal slice of tensor $\mathcal{X}$
$\mathcal{X}_f = \text{fft}(\mathcal{X}, [], 3)$	the fast Fourier transformation (FFT)
n, t, m	the number of samples, anchors, and views
d^v	the dimension of v-th view
$\mathbf{X}^v \in \mathbb{R}^{n \times d^v}$	the feature matrix of v-th view
$\mathbf{A} \in \mathbb{R}^{t \times l}$	the anchor matrix with l dimension
$\mathbf{Z}^v, \mathbf{Z}^v_{anc}$	the representation matrix in v-th view
$\mathcal{Z}, \mathcal{Z}_{anc}$	the representation tensor
$\mathbf{E}^v \in \mathbb{R}^{n \times d^v}$	the reconstruction error in v-th view
$\|\cdot\|_F$, $\|\cdot\|_{2,1}$	the Frobenius norm, $\ell_{2,1}$ norm
$\|\cdot\|_{\circledast}$	the nuclear norm

Then, given the tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, we will introduce some operations related to tensor [40].

- Transposition of tensor $\mathcal{X}^T \in \mathbb{R}^{n_2 \times n_1 \times n_3}$, which means that each frontal slice of the tensor is transposed.
- Cyclic expansion of the tensor $circ(\mathcal{X}) \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}$:

$$circ(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^1 & \mathcal{X}^{n_3} & \cdots & \mathcal{X}^2 \\ \mathcal{X}^2 & \mathcal{X}^1 & \cdots & \mathcal{X}^3 \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{n_3} & \mathcal{X}^{n_3-1} & \cdots & \mathcal{X}^1 \end{bmatrix}. \tag{1}$$

- Tensor unfolding and folding:

$$\begin{aligned} &unfold(\mathcal{X}) = \left[\mathcal{X}^1, \mathcal{X}^2, \cdots, \mathcal{X}^{n_3}\right]^T \in \mathbb{R}^{n_1 n_3 \times n_2}. \\ &\mathcal{X} = fold(unfold(\mathcal{X})). \end{aligned} \tag{2}$$

- t-product $\mathcal{X} * \mathcal{Y} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$:

$$\mathcal{X} * \mathcal{Y} = fold(circ(\mathcal{X}) \cdot unfold(\mathcal{Y})). \tag{3}$$

- Orthogonal tensor: The tensor $\mathcal{X}$ is orthogonal if $\mathcal{X}^T * \mathcal{X} = \mathcal{X} * \mathcal{X}^T = \mathcal{I}$, where $\mathcal{I} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is identity tensor whose first frontal slice is $\mathcal{I}^1 = \mathbf{I} \in \mathbb{R}^{n_1 \times\ n_1}$, and the other frontal slices are $\mathcal{I}^k = \mathbf{0}, \forall k = 2, 3, \ldots, n_3$.

Based on the above tensor operations, we can give the definition of tensor Singular Value Decomposition (t-SVD) and Tensor Nuclear Norm (TNN).

Definition 1 *(t-SVD) [41] Given tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then the t-SVD of $\mathcal{X}$ is:*

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T \tag{4}$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f-diagonal tensor.

Definition 2 *(TNN) [31]–[33] Given a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the nuclear norm of the tensor is defined as:*

$$\|\mathcal{X}\|_{\circledast} = \frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{i=1}^{h} \mathcal{S}_f^k(i, i) \tag{5}$$

where $h = \min(n_1, n_2)$ and $\mathcal{S}_f$ is from the t-SVD of $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ in the Fourier domain.

2.2 Multi-view Subspace Clustering

Multi-view subspace clustering [15], [17], [42] has recently attracted more attention due to its ability to handle high-dimensional data efficiently. It usually adopts the self-representation strategy to learn the multi-view representations, then an affinity matrix can be fused and fed into spectral clustering to obtain the clustering result. Given a multi-view data $\{\mathbf{X}^v\}_{v=1}^m$ with n samples and m views, the general framework of multi-view subspace clustering algorithms can be expressed as follows:

$$\begin{aligned} &\min_{\{\mathbf{Z}^v\}_{v=1}^m, \{\mathbf{E}^v\}_{v=1}^m} \Omega(\{\mathbf{E}^v\}_{v=1}^m) + \beta \mathcal{R}(\{\mathbf{Z}^v\}_{v=1}^m) \\ &s.t.\ \forall v, \mathbf{X}^v = \mathbf{Z}^v \mathbf{X}^v + \mathbf{E}^v, \end{aligned} \tag{6}$$

where $\mathbf{X}^v \in \mathbb{R}^{n \times d^v}$, d^v is the dimension of v-th view. $\Omega(\cdot)$ denotes the loss function. $\mathcal{R}(\cdot)$ means some regularization operations on the representation matrices. Many multi-view clustering methods can be obtained by combining different regularizations and loss functions. For instance, Cao *et al.* [16] adopt the Hilbert Schmidt Independence Criterion (HSIC) to diversify the multi-view representations. Chen *et al.* [43] propose smooth regularization and cross-view consistency regularization to explicitly preserve the locality and exploit the subspace grouping effect, respectively. These algorithms constrain the representation matrices in individual views or between pairs of views, resulting in the absence of global information about all views. Recently, some works are gradually adopting global constraint (*i.e.*, low-rank tensor constraint) to enhance the exploration of high-order correlation among views, called tensorial multi-view subspace clustering. Specifically, given a multi-view data $\{\mathbf{X}^v\}_{v=1}^m$ with n samples and m views, the general framework of tensorial multi-view subspace clustering algorithms can be expressed as follows:

$$\begin{aligned} &\min_{\mathcal{Z}, \{\mathbf{E}^v\}_{v=1}^m} \mathcal{L}\mathcal{R}(\mathcal{Z}) + \alpha \Omega(\{\mathbf{E}^v\}_{v=1}^m) + \beta \mathcal{R}(\{\mathbf{Z}^v\}_{v=1}^m) \\ &s.t.\ \forall v, \mathbf{X}^v = \mathbf{Z}^v \mathbf{X}^v + \mathbf{E}^v, \mathcal{Z} = \Phi(\mathbf{Z}^1, \mathbf{Z}^2, \cdots, \mathbf{Z}^m), \end{aligned} \tag{7}$$

where α, β are the trade-off parameters. $\Phi(\cdot)$ denotes merging and rotating operation [26], which merges anchor-representation matrices $\{\mathbf{Z}^v\}_{v=1}^m$ to a three-order tensor $\mathcal{Z}$ with dimension $n \times m \times n$. $\mathcal{L}\mathcal{R}(\cdot)$ denotes the low-rank constrain of tensor (*e.g.*, TNN [44], TAR [36], TLS_pN [35], etc.). $\Omega(\cdot)$ measures the reconstruction error of self-representation typically using the matrix Frobeniusm norm or the $\ell_{2,1}$ norm. $\mathcal{R}(\cdot)$ means some specific constraint terms for the representation matrices, for example, the hyper-Laplacian regularizer [45] and the sliced sparse gradient regularization [36]. Once the optimal representation matrices $\{\mathbf{Z}^v\}_{v=1}^m$ have been learned, the fused affinity matrix

$\mathbf{S} = \frac{1}{m}\sum_{v=1}^{m}(|\mathbf{Z}^v| + |\mathbf{Z}^{vT}|)$ can be fed into the spectral clustering to obtain the final clustering result. Although the low-rank tensor constraint significantly enhances the learning of global information among views, it also introduces several issues such as severe computational burden, tensor-rank approximation problem, and consistency-complementarity balancing problem. To the best of our knowledge, no tensorial multi-view subspace clustering algorithm can simultaneously address all of these problems.

3 PROPOSED METHOD

3.1 Efficient and Scalable Tensorial Multi-View Clustering Framework (ESTMC)

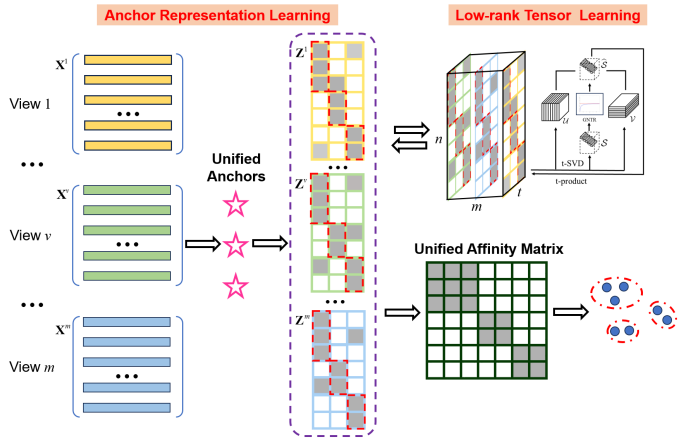


Fig. 1: Framework of the proposed ESTMC. Given multi-view data $\{\mathbf{X}^v\}_{v=1}^m$ with m views, we first learn a set of anchors in a unified space to construct the representation matrices $\{\mathbf{Z}^v\}_{v=1}^m$. Then we adopt the low-rank tensor constraint based on non-convex functions to learn the high-order correlation hidden in representation matrices.

Tensorial multi-view subspace clustering methods can explore the high-order correlation among views to achieve satisfactory performance, which benefits from the t-SVD-based low-rank tensor learning framework. As shown in Fig. 2, the t-SVD-based low-rank tensor learning framework can use the low-rank property of the tensor as a view-level constraint to significantly improve the consistency of multi-view representation matrices, leading to a compact affinity matrix that shows a clear clustering structure. Although such algorithms have made impressive progress in the field of multi-view clustering, there are still two key issues to be addressed. The first problem is the tensor rank approximation, where the learning of the tensor low-rank structure directly affects the clustering structure of the affinity matrix. Second, due to the extremely high computational complexity of tensor correlation operations (t-SVD and FFT), the efficiency of processing large-scale datasets will be seriously affected.

For the tensor rank approximation problem, it is equivalent to the traditional matrix rank approximation problem. For example, the commonly used Tensor Nuclear Norm (TNN) is an extension of the traditional nuclear norm, the least convex envelope of the rank function. However, some works [35], [36] have proved that TNN is a loose and biased

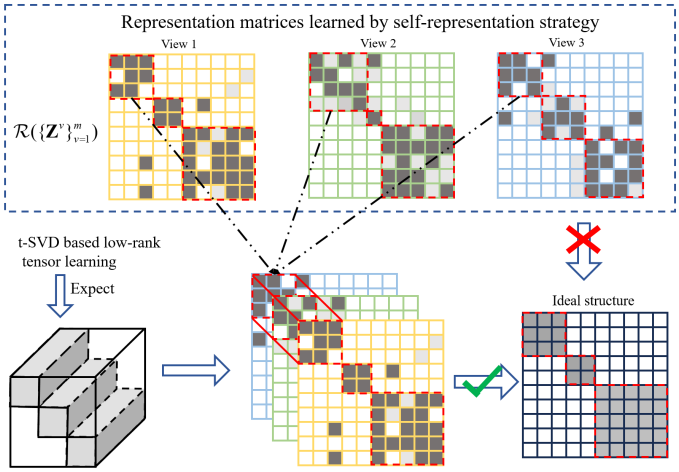


Fig. 2: Illustration of the t-SVD-based low-rank tensor learning framework. The consistent diagonal block structure can be efficiently captured through the low-rank property of the tensor, whereas traditional fusion methods at the matrix level (weighted fusion strategy, averaged fusion strategy) will confuse diagonal block structures.

approximation, which under-penalizes the smaller singular values and over-penalizes the larger singular values. Usually, the large singular values represent the main information, and small singular values indicate noise, thus TNN often results in noise residuals that disrupt the clustering structure. This motivates us to approximate the rank function by non-convex surrogate functions that have a better approximation to the rank function and provide appropriate penalties for varying singular values. So, the tensor rank based on the non-convex surrogate function can explore a better low-rank property of the representation tensor than TNN.

To this end, we design a standard form of tensor rank based on the non-convex surrogate function, called Generalized Non-convex Tensor Rank (GNTR) (See Definition 3), which can extend most existing non-convex surrogate functions (Table 2) or their variants to TMC framework.

In the next section, we also propose an improved Enhanced Tensor Rank (ETR) that incorporates the strengths of several existing non-convex functions.

Definition 3 *Given a tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then the Generalized Non-convex Tensor Rank (GNTR) is defined as:*

$$\begin{aligned}\|\mathcal{Z}\|_{GNTR} &= \frac{1}{n_3}\sum_{k=1}^{n_3}\left\|\mathcal{Z}_f^k\right\|_{GNTR} \\ &= \frac{1}{n_3}\sum_{k=1}^{n_3}\sum_{i=1}^{h}\psi\left(\mathcal{S}_f^k(i,i),\delta\right),\end{aligned} \tag{8}$$

where δ denotes a set of parameters, $h = \min(n_1, n_2)$ and $\mathcal{S}_f$ is obtained by t-SVD of $\mathcal{Z}_f = \mathcal{U}_f\mathcal{S}_f\mathcal{V}_f^T$ in Fourier domain.

The $\psi(x,\delta)$ represents the non-convex surrogate functions, which satisfies the following assumptions:

1) *$\psi(x,\delta)$: $\mathbb{R} \Rightarrow \mathbb{R}^+$ is continuous, concave and monotonically increasing on $[0,\infty)$. It is possibly non-smooth.*
2) *$x = 0 \Rightarrow \psi(0,\delta) = 0$.*

For the computational complexity problem, constructing a space-saving representation tensor is an effective way to

TABLE 2: Popular non-convex surrogate functions.

Name	Formula $\psi(x,\delta)$ (δ is parameter)
ℓ_δ-norm [46]	$\frac{(1+\delta)x}{\delta+x}$
Arctan [47]	$\text{Arctan}\frac{\lvert x\rvert}{\delta}$
Geman [48]	$\frac{\delta x}{x+\delta}$
Laplace [49], [50]	$1-exp(-\frac{x}{\delta})$
Logdet [51], [52]	$\log(1+x^2)$

alleviate the computational burden of tensor-related operations (e.g., t-SVD and FFT). Inspired by the anchor-based approaches [21], [53], we adopt the anchor representation strategy, which selects several representative samples as anchor dictionary to characterize all samples,

$$\forall v,\ \mathbf{X}^v = \mathbf{Z}^v_{anc}\mathbf{A}\mathbf{P}^v + \mathbf{E}^v \tag{9}$$

where $\mathbf{A} \in \mathbb{R}^{t\times l}$ is the anchors in unified anchor space with l dimensions, and t is the number of anchors. $\mathbf{P}^v \in \mathbb{R}^{l\times d_v}$ denotes the space mapping matrix of v-th view. $\mathbf{Z}^v_{anc} \in \mathbb{R}^{n\times t}$ is the anchor-representation matrix of v-th view. Apparently, the constructed anchor-representation tensor $\mathcal{Z}_{anc} = \Phi(\mathbf{Z}^1_{anc}, ..., \mathbf{Z}^m_{anc}) \in \mathbb{R}^{t\times m\times n}$ is space-saving and will take less computational complexity on tensor-related operations. Therefore, the objective function of our proposed Efficient and Scalable Tensorial Multi-View Clustering Framework (ESTMC) is as follows:

$$\begin{aligned}
&\min_{\mathcal{Z}_{anc},\mathbf{E},\mathbf{A},\{\mathbf{P}^v\}_{v=1}^m} \|\mathcal{Z}_{anc}\|_{GNTR} + \lambda\|\mathbf{E}\|_{2,1}\\
&s.t.\ \mathbf{X}^v = \mathbf{Z}^v_{anc}\mathbf{A}\mathbf{P}^v + \mathbf{E}^v, v = 1,2,\ldots,m,\\
&\quad \mathcal{Z}_{anc} = \Phi(\mathbf{Z}^1_{anc}, \mathbf{Z}^2_{anc}, \cdots, \mathbf{Z}^m_{anc}),\\
&\quad \mathbf{E} = [\mathbf{E}^1, \mathbf{E}^2, \cdots, \mathbf{E}^m]^T, \mathbf{A}\mathbf{A}^T = \mathbf{I}, \mathbf{P}^v\mathbf{P}^{vT} = \mathbf{I},
\end{aligned} \tag{10}$$

where γ is the parameter. $\Phi(\cdot)$ denotes merging and rotating operation [26], which merges anchor-representation matrices $\{\mathbf{Z}^v_{anc}\}_{v=1}^m$ to a three-order tensor $\mathcal{Z}_{anc}$ with dimension $t\times m\times n$.

Once the optimal representation matrices $\{\mathbf{Z}^v_{anc}\}_{v=1}^m$ are obtained, the unified affinity matrix can be fused as $\mathbf{S_Z} = \frac{1}{m}\sum_{v=1}^m \mathbf{Z}^v_{anc}{\mathbf{Z}^v_{anc}}^T \in \mathbb{R}^{n\times n}$. Inspired by [20], the spectral embedding $\mathbf{F} \in \mathbb{R}^{n\times c}$ can be obtained by directly imposing the Singular Value Decomposition (SVD) on $\hat{\mathbf{Z}} = \frac{1}{\sqrt{m}}[\mathbf{Z}^1_{anc}, \cdots, \mathbf{Z}^m_{anc}] \in \mathbb{R}^{n\times mt}$. According to Theorem 1, it has linear time complexity $\mathcal{O}(nm^2t^2)$.

Theorem 1 *The left singular vectors of* $\hat{\mathbf{Z}} = \frac{1}{\sqrt{m}}[\mathbf{Z}^1, \cdots, \mathbf{Z}^m] \in \mathbb{R}^{n\times mt}$ *is the same as the eigenvectors of* $\mathbf{S} = \hat{\mathbf{Z}}\hat{\mathbf{Z}}^T$.

Proof The singular value decomposition (SVD) of $\hat{\mathbf{Z}}$ is indicated as $\mathbf{U}\Sigma\mathbf{V}^T$, then

$$\mathbf{S} = \hat{\mathbf{Z}}\hat{\mathbf{Z}}^T = (\mathbf{U}\Sigma\mathbf{V}^T)(\mathbf{U}\Sigma\mathbf{V}^T)^T = \mathbf{U}\Sigma^2\mathbf{U}^T. \tag{11}$$

Therefore, the left singular vectors of $\hat{\mathbf{Z}}$ are the same as the eigenvectors of $\mathbf{S}$. □

Inspired by the alternation direction method of multipliers (ADMM) [54], we introduce the auxiliary tensor variable $\mathcal{G}$, so the model (10) can be rewritten as the following unconstrained problem,

$$\begin{aligned}
&\mathcal{L}(\{\mathbf{Z}^v_{anc}\}_{v=1}^m, \mathbf{A}, \{\mathbf{P}^v\}_{v=1}^m, \mathcal{G}, \mathbf{E}, \{\mathbf{Y}^v\}_{v=1}^m, \mathcal{W})\\
&= \|\mathcal{G}\|_{GNTR} + \lambda\|\mathbf{E}\|_{2,1} + \langle\mathcal{W}, \mathcal{Z}_{anc} - \mathcal{G}\rangle\\
&+ \sum_{v=1}^m \big(\langle\mathbf{Y}^v, \mathbf{X}^v - \mathbf{Z}^v_{anc}\mathbf{A}\mathbf{P}^v - \mathbf{E}^v\rangle\\
&+ \frac{\mu}{2}\|\mathbf{X}^v - \mathbf{Z}^v_{anc}\mathbf{A}\mathbf{P}^v - \mathbf{E}^v\|_F^2\big) + \frac{\rho}{2}\|\mathcal{Z}_{anc} - \mathcal{G}\|_F^2,
\end{aligned} \tag{12}$$

where $\mathcal{W}$ and $\{\mathbf{Y}^v\}_{v=1}^m$ are Lagrange multipliers, μ and ρ are penalty parameters to control convergence. Then, we solve the variables in Eq. (12) through the following five subproblems.

•$\{\mathbf{Z}^v_{anc}\}_{v=1}^m$**-Subproblem** Fixing the other variables leads to the following problem for $\mathbf{Z}^v_{anc}$,

$$\begin{aligned}
\arg\min_{\mathbf{Z}^v_{anc}}\ & Tr(\mathbf{W}^{vT}(\mathbf{Z}^v_{anc} - \mathbf{G}^v)) + \frac{\rho}{2}\|\mathbf{Z}^v_{anc} - \mathbf{G}^v\|_F^2\\
&+ \langle\mathbf{Y}^v, \mathbf{X}^v - \mathbf{Z}^v_{anc}\mathbf{A}\mathbf{P}^v - \mathbf{E}^v\rangle\\
&+ \frac{\mu}{2}\|\mathbf{X}^v - \mathbf{Z}^v_{anc}\mathbf{A}\mathbf{P}^v - \mathbf{E}^v\|_F^2.
\end{aligned} \tag{13}$$

this problem is a quadratic convex optimization problem, so derivative for $\mathbf{Z}^v$ and equals to 0 :

$$\begin{aligned}
\mathbf{Z}^v_{anc} = \Big(&\mathbf{Y}^v\mathbf{P}^{vT}\mathbf{A}^T + \mu\mathbf{X}^v\mathbf{P}^{vT}\mathbf{A}^T + \rho\mathbf{G}^v - \mathbf{W}^v\\
&-\mu\mathbf{E}^v\mathbf{P}^{vT}\mathbf{A}^T\Big)\left((\rho+\mu)\mathbf{I}\right)^{-1}
\end{aligned} \tag{14}$$

•**E-Subproblem** Fixing the other variables, the problem with $\mathbf{E}$ is formulated as,

$$\arg\min_{\mathbf{E}} \frac{\lambda}{\mu}\|\mathbf{E}\|_{2,1} + \frac{1}{2}\left\|\mathbf{E} - \hat{\mathbf{E}}\right\|_F^2, \tag{15}$$

where $\hat{\mathbf{E}} = [\mathbf{X}^1 - \mathbf{Z}^1_{anc}\mathbf{A}\mathbf{P}^1 + \frac{1}{\mu}\mathbf{Y}^1, \cdots, \mathbf{X}^m - \mathbf{Z}^m_{anc}\mathbf{A}\mathbf{P}^m + \frac{1}{\mu}\mathbf{Y}^m]^T$. Its solution can be obtained by $\ell_{2,1}$ minimization thresholding operator as in [55],

$$\mathbf{E}_{:,j} = \begin{cases} \dfrac{\left\|\hat{\mathbf{E}}_{:,j}\right\|_2 - \frac{\lambda}{\mu}}{\left\|\hat{\mathbf{E}}_{:,j}\right\|_2}\hat{\mathbf{E}}_{:,j}, & \left\|\hat{\mathbf{E}}_{:,j}\right\|_2 > \dfrac{\lambda}{\mu},\\ 0, & \text{otherwise.}\end{cases} \tag{16}$$

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

•$\mathcal{G}$**-Subproblem** When other variables are fixed, the subproblem for $\mathcal{G}$ is formulated as,

$$\arg\min_{\mathcal{G}} \frac{1}{\rho}\|\mathcal{G}\|_{GNTR} + \frac{1}{2}\left\|\mathcal{G} - (\mathcal{Z}_{anc} + \frac{\mathcal{W}}{\rho})\right\|_F^2. \tag{17}$$

We refer to this problem as the Generalized Nonconve Tensor Rank Minimization problem (GNTRM), which can be solved by the following theorem.

Theorem 2 *Suppose* $\mathcal{G}, \mathcal{Q} \in \mathbb{R}^{n_1\times n_2\times n_3}$ *with t-SVD* $\mathcal{Q} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ *and* $\beta > 0$, *the Generalized Non-convex Tensor Rank Minimization problem (GNTRM) can be described as follows,*

$$\arg\min_{\mathcal{G}} \beta\|\mathcal{G}\|_{GNTR} + \frac{1}{2}\|\mathcal{G} - \mathcal{Q}\|_F^2. \tag{18}$$

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

$$\mathcal{G}^* = \mathcal{U} * ifft(Prox_{f,\beta}(\mathcal{S}_f), [], 3) * \mathcal{V}^T, \tag{19}$$

Algorithm 1 Optimization Algorithm of ESTMC

Input: Multi-view data matrix $\{\mathbf{X}^1, \ldots, \mathbf{X}^m\}$, cluster number c, trade-off parameter λ and anchor number t.
Output: Clustering results $\mathbf{C}$.
Initialize: $\forall v, \mathbf{Z}^v = \mathbf{0}, \mathbf{E}^v = \mathbf{0}, \mathbf{Y}^v = \mathbf{0}, \mathcal{G} = \mathbf{0}, \mathcal{W} = \mathbf{0}$, $\mu = 10^{-5}, \rho = 10^{-4}, \eta_\mu = \eta_\rho = 2, \mu_{max} = \rho_{max} = 10^{10}, \epsilon = 10^{-7}$.
while not converge **do**
  Update $\mathbf{Z}^v_{anc}, (v = 1, 2, \ldots, m)$ by Eq.(14);
  Update $\mathbf{E}$ by Eq.(16);
  Update $\mathbf{P}^v, (v = 1, 2, \ldots, m)$ by Eq.(28);
  Update $\mathbf{A}$ by Eq.(29);
  Update $\mathbf{Y}^v, (v = 1, 2, \ldots, m)$ by Eq.(30);
  Update $\mathcal{W}$ by Eq.(30);
  Update parameters μ and ρ by Eq.(30);
  Check the convergence conditions: $\|\mathbf{X}^v - \mathbf{Z}^v_{anc}\mathbf{A}\mathbf{P}^v - \mathbf{E}^v\|_\infty < \epsilon$ & $\|\mathcal{Z}^v_{anc} - \mathcal{G}^v\|_\infty < \epsilon$;
end while
Output clustering results $\mathbf{C}$ via performing k-means on $\mathbf{F}$.

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

$$Prox_{f,\beta}(\mathcal{S}_f^k(i,i)) = \arg\min_{x \geq 0} \frac{1}{2}(x - \mathcal{S}_f^k(i,i))^2 + \beta f(x), \quad (20)$$

where $f(x) = \psi(x, \delta)$.

Theorem 2 can be proved with the following lemma.

Lemma 1 *[46] Given matrices $\mathbf{G}, \mathbf{Q} \in \mathbb{R}^{n_1 \times n_2}$, and $\mathbf{Q} = \mathbf{U}\mathbf{S}_Q\mathbf{V}^T$ is the SVD of $\mathbf{Q}$ and $\beta > 0$, then an optimal solution to the following problem*

$$\min_{\mathbf{G}} \beta \|\mathbf{G}\|_{GNTR} + \frac{1}{2}\|\mathbf{G} - \mathbf{Q}\|_F^2, \quad (21)$$

is $\mathbf{G}^* = \mathbf{U}\mathbf{S}_G^*\mathbf{V}^T$, where $\mathbf{S}_G^* = diag(\sigma^*)$ and $\sigma^* = Prox_{f,\beta}(\sigma_Q)$. And $Prox_{f,\beta}(\sigma_Q)$ is the Moreau-Yosida operator [56] defined as

$$Prox_{f,\beta}(\sigma_Q) := \arg\min_{\sigma \geq 0} \beta f(\sigma) + \frac{1}{2}\|\sigma - \sigma_Q\|_2^2, \quad (22)$$

where $f(x) = \psi(x, \delta)$.

Proof In Fourier domain, based on the fact that $\|\mathcal{X}\|_F^2 = \frac{1}{n_3}\|\mathcal{X}_f\|_F^2$, the objective function $\frac{1}{2}\|\mathcal{G} - \mathcal{Q}\|_F^2 + \beta\|\mathcal{G}\|_{GNTR}$ can be transformed to the following form,

$$\begin{aligned} &\frac{1}{2}\|\mathcal{G} - \mathcal{Q}\|_F^2 + \beta\|\mathcal{G}\|_{GNTR} \\ =&\frac{1}{2n_3}\|\mathcal{G}_f - \mathcal{Q}_f\|_F^2 + \frac{\beta}{n_3}\sum_{k=1}^{n_3}\left\|\mathcal{G}_f^k\right\|_{GNTR} \\ =&\frac{1}{n_3}\sum_{k=1}^{n_3}\left(\frac{1}{2}\left\|\mathcal{G}_f^k - \mathcal{Q}_f^k\right\|_F^2 + \beta\left\|\mathcal{G}_f^k\right\|_{GNTR}\right) \end{aligned} \quad (23)$$

As a result, the original tensor optimization problem can be reformed into the following n_3 independent matrix optimization problem:

$$\arg\min_{\mathcal{G}_f^k} \frac{1}{2}\left\|\mathcal{G}_f^k - \mathcal{Q}_f^k\right\|_F^2 + \beta\left\|\mathcal{G}_f^k\right\|_{GNTR}, \quad (24)$$

for $1 \leq k \leq n_3$.

Here, the SVD of $\mathcal{Q}_f^k$ is $\mathcal{Q}_f^k = \mathcal{U}_f^k\mathcal{S}_f^k(\mathcal{V}_f^k)^H$. According to Lemma 1, the optimal solution of (24) is

$$\mathcal{G}_f^{*k} = \mathcal{U}_f^k Prox_{f,\beta}(\mathcal{S}_f^k)(\mathcal{V}_f^k)^H, \quad (25)$$

where $Prox_{f,\beta}(\mathcal{S}_f^k(i,i))$ is given by solving the following problem:

$$Prox_{f,\beta}(\mathcal{S}_f^k(i,i)) = \arg\min_{x \geq 0} \frac{1}{2}(x - \mathcal{S}_f^k(i,i))^2 + \beta f(x) \quad (26)$$

where $f(x) = \psi(x, \delta)$. □

Eq. (26) is a combination of concave and convex functions, so we can use the Difference of Convex (DC) programming [57] to acquire a closed-form solution,

$$\tau^{iter+1} = \left(\mathcal{S}_f^k(i,i) - \frac{\partial f(\tau^{iter})}{\rho}\right)_+, \quad (27)$$

where $\tau = Prox_{f,\beta}(\mathcal{S}_f^k(i,i)),\ f(x) = \psi(x, \delta)$ and $iter$ is the number of iterations.

•$\{\mathbf{P}^v\}_{v=1}^m$**-Subproblem** Fixing the other variables, $\mathbf{P}^v$ can be updated by,

$$\mathbf{P}^{v*} = \arg\max_{\mathbf{P}^v\mathbf{P}^{vT} = \mathbf{I}} Tr(\mathbf{P}^{vT}\mathbf{N}^v), \quad (28)$$

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

•**A-Subproblem** Fixing the other variables, $\mathbf{A}$ can be updated by,

$$\mathbf{A} = \arg\max_{\mathbf{A}\mathbf{A}^T = \mathbf{I}} Tr(\mathbf{A}^T\mathbf{M}), \quad (29)$$

where $\mathbf{M} = \sum_{v=1}^m {\mathbf{Z}^v_{com}}^T(\mu\mathbf{X}^v + \mathbf{Y}^v - \mu\mathbf{E}^v)\mathbf{P}^{vT}$. The optimal solution of $\mathbf{A}$ is $\mathbf{U}_A{\mathbf{V}_A}^T$, where $\mathbf{U}_A$ and $\mathbf{V}_A$ are the left and right singular matrix of $\mathbf{M}$.

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

$$\begin{cases} \mathbf{Y}^v = \mathbf{Y}^v + \mu(\mathbf{X}^v - \mathbf{Z}^v_{anc}\mathbf{A}\mathbf{P}^v - \mathbf{E}^v) \\ \mathcal{W} = \mathcal{W} + \rho(\mathcal{Z}_{anc} - \mathcal{G}) \\ \mu = \eta_\mu\mu,\ \mu = \min(\mu, \mu_{\max}) \\ \rho = \eta_\rho\rho,\ \rho = \min(\eta, \rho_{\max}) \end{cases} \quad (30)$$

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

3.2 Generic to Specific: Compromise Between Consistency and Complementarity

In this section, we further discuss how to explicitly capture complementary information among views based on the proposed ESTMC framework. Usually, the t-SVD-based low-rank tensor learning framework focuses on capturing the consistency in different views and discards some view-specific information. However, samples in similar classes or at the class boundary tend to exhibit similar representations in most views. This may lead to incorrect classification when relying solely on the captured consistent information. To obtain more discriminative representations for these samples, we strive to leverage consistent and complementary

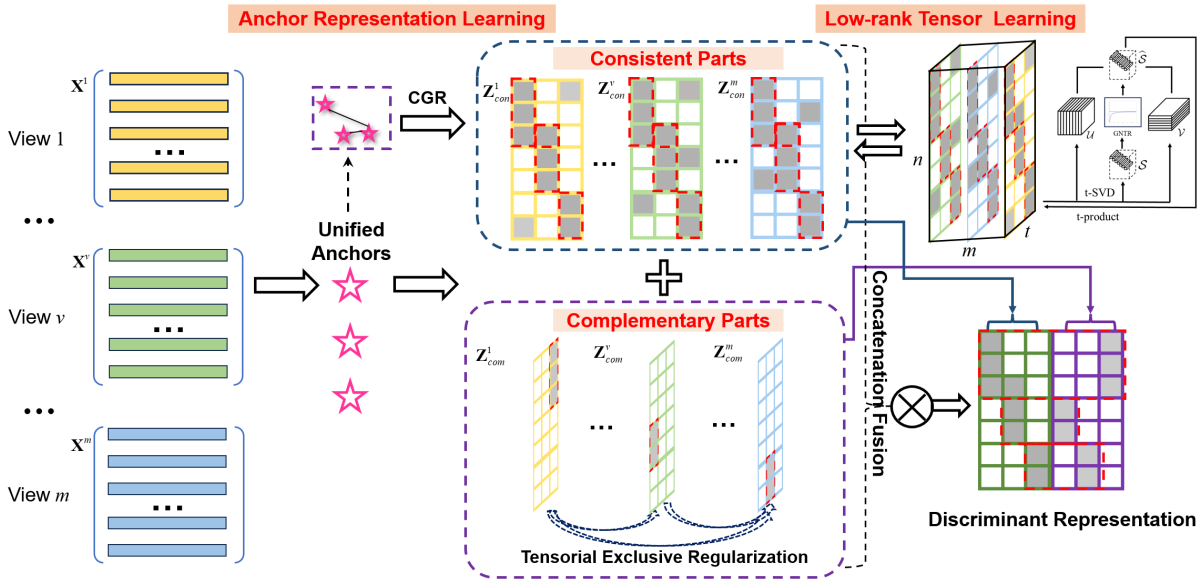


Fig. 3: Framework of the proposed ESTMC-C^2. Given multi-view data $\{\mathbf{X}^v\}_{v=1}^m$ with m views, we first adopt the learned anchors to construct two groups of representation matrices $\{\mathbf{Z}_{con}^v\}_{v=1}^m$, $\{\mathbf{Z}_{com}^v\}_{v=1}^m$, which denotes the consistent parts and complementary parts, respectively. Then we adopt the non-convex function-based low-rank tensor constraint and CGR on consistent parts $\{\mathbf{Z}_{con}^v\}_{v=1}^m$ to explore the high-order correlation and consistent structure. Meanwhile, the Tensor Exclusive Regularization is imposed on complementary parts $\{\mathbf{Z}_{com}^v\}_{v=1}^m$ to capture view-specific information. We design a concatenation-fusion strategy to obtain a more discriminate representation for subsequent clustering.

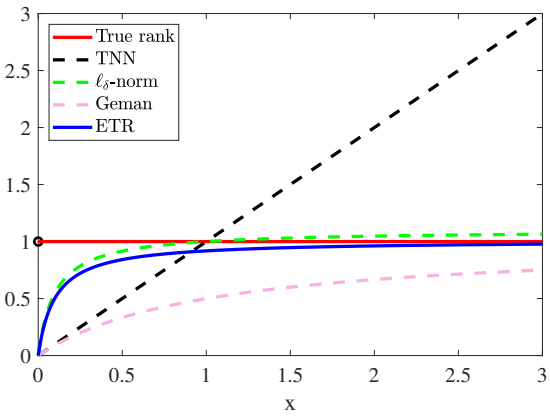


Fig. 4: Comparison of different approximation functions with the same parameter.

information from multiple views simultaneously. Specifically, two types of representation matrices $\{\mathbf{Z}_{con}^v\}_{v=1}^m$ and $\{\mathbf{Z}_{com}^v\}_{v=1}^m$ are obtained via

$$\forall v,\ \mathbf{X}^v = (\mathbf{Z}_{con}^v + \mathbf{Z}_{com}^v)\mathbf{A}\mathbf{P}^v + \mathbf{E}^v, \tag{31}$$

where $\mathbf{Z}_{con}^v \in \mathbb{R}^{n\times t}$ and $\mathbf{Z}_{com}^v \in \mathbb{R}^{n\times t}$ denote the consistent and complementary parts in v-th view, respectively.

For the complementary parts, similar to [58], we introduce the Tensorial Exclusive Regularization (TER), defined in Definition 4, to explicitly capture view-specific information.

Definition 4 *Given a tensor $\mathcal{Z} \in \mathbb{R}^{n_1\times n_2\times n_3}$, then the Tensorial Exclusive Regularization (TER) is defined as:*

$$\|\mathcal{Z}\|_{TER} = \frac{1}{n_3}\sum_{v=1}^{n_3}\sum_{w=1}^{n_3}\langle \mathbf{Z}^v, \mathbf{Z}^w\rangle \tag{32}$$

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

For the consistent parts, based on the works [46], [48], we propose a novel Enhanced Tensor Rank (ETR) with an improved non-convex surrogate function $f_{ETR}(x) = e^{\delta^2}x/(\delta+x)$ instead of existing functions. ETR aims to make a compromise between ℓ_δ-norm and the Geman function. From Fig. 4, it is clear to see that ETR encourages the penalty for small singular values to be increased and the penalty for large singular values to be decreased. According to the Definition 3, the definition of ETR is shown as follows.

Definition 5 *[39] 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:*

$$\begin{aligned}\|\mathcal{Z}\|_{ETR} &= \frac{1}{n_3}\sum_{k=1}^{n_3}\left\|\mathcal{Z}_f^k\right\|_{ETR} \\ &= \frac{1}{n_3}\sum_{k=1}^{n_3}\sum_{i=1}^{h}\left(\frac{e^{\delta^2}\mathcal{S}_f^k(i,i)}{\delta+\mathcal{S}_f^k(i,i)}\right),\end{aligned} \tag{33}$$

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

Since the representation matrices are generated through dictionary learning techniques (such as anchor representation, self-representation, and matrix factorization), the learning of consistent non-linear structure is typically absent. To mitigate this, graph-based structural constraints are typically applied. However, such constraint usually requires constructing an adjacency matrix in sample space, often with a high complexity $\mathcal{O}(n^2)$. To improve efficiency, we use the anchor space to construct a resource-saving constraint, called Consistent Geometric Regularization (CGR), which is defined as Definition 6.

Definition 6 *Given an anchor-representation tensor* $\mathcal{Z} = \Phi(\mathbf{Z}^1, \cdots, \mathbf{Z}^m) \in \mathbb{R}^{t\times m\times n}$ *and unified anchors* $\mathbf{A}$*, then the Consistent Geometric Regularization (CGR) is defined as:*

$$\|\mathcal{Z}\|_{CGR} = \sum_{v=1}^{m} \|\mathbf{Z}^v\|_{CGR} = \sum_{v=1}^{m} Tr(\mathbf{Z}^v \mathbf{L} (\mathbf{Z}^v)^T), \tag{34}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{B}$ *is the Laplacian matrix of* $\mathbf{B} \in \mathbb{R}^{t\times t}$*, which is the adjacency matrix of unified anchors* $\mathbf{A} = [\boldsymbol{a}_1; \cdots; \boldsymbol{a}_t]$ *and is constructed by solving following problem:*

$$\min_{\boldsymbol{b}_i \mathbf{1} = \mathbf{1}, \boldsymbol{b}_i \geq 0} \sum_{i,j=1}^{t} \|\boldsymbol{a}_i - \boldsymbol{a}_j\|_2^2 b_{ij} + \omega \sum_{i=1}^{t} \|\boldsymbol{b}_i\|_2^2, \tag{35}$$

the closed-form solution can be obtained from [8]. The degree matrix $\mathbf{D}$ *is a diagonal matrix whose i-th diagonal element is computed by* $\mathbf{D}(i,i) = \sum_{j=1}^{t} \mathbf{B}(i,j)$*.*

By integrating Eq. (10), Eq. (31), Eq. (34), and Eq. (32), the ESTMC can be further boosted to ESTMC-C^2 with the following objective function:

$$\begin{aligned} \min_{\mathbf{A}, \mathcal{Z}_{con}, \mathcal{Z}_{com}, \mathbf{E}, \{\mathbf{P}^v\}_{v=1}^m} & \|\mathcal{Z}_{con}\|_{ETR} + \lambda \|\mathbf{E}\|_{2,1} \\ & + \beta \|\mathcal{Z}_{con}\|_{CGR} + \alpha \|\mathcal{Z}_{com}\|_{TER} \\ s.t.\ & \mathbf{X}^v = (\mathbf{Z}_{con}^v + \mathbf{Z}_{com}^v)\mathbf{A}\mathbf{P}^v + \mathbf{E}^v, v = 1, 2, \ldots, m, \\ & \mathcal{Z}_{con} = \Phi(\mathbf{Z}_{con}^1, \cdots, \mathbf{Z}_{con}^m), \mathcal{Z}_{com} = \Psi(\mathbf{Z}_{com}^1, \cdots, \mathbf{Z}_{com}^m), \\ & \mathbf{E} = [\mathbf{E}^1, \mathbf{E}^2, \cdots; \mathbf{E}^m]^T, \mathbf{A}\mathbf{A}^T = \mathbf{I}, \mathbf{P}^v \mathbf{P}^{vT} = \mathbf{I}. \end{aligned} \tag{36}$$

where λ, β, and α are trade-off parameters. $\Phi(\cdot)$ denotes merging and rotating operation, which merges anchor-representation matrices $\{\mathbf{Z}_{con}^v\}_{v=1}^m$ to a three-order tensor $\mathcal{Z}_{con}$ with dimension of $t \times m \times n$. $\Psi(\cdot)$ denotes merging operation that merges $\{\mathbf{Z}_{com}^v\}_{v=1}^m$ into three-order tensor with a dimension of $n \times t \times m$.

Then, rather than using only the consistent parts for subsequent clustering task [59], we employ a concatenation-fusion strategy to combine both the learned consistent parts and complementary parts to obtain more discriminate representations $\mathbf{U}^v = [\mathbf{Z}_{con}^v, \mathbf{Z}_{com}^v] \in \mathbb{R}^{n\times 2t}$. Then similar to ESTMC, the spectral embedding $\mathbf{F} \in \mathbb{R}^{n\times c}$ can be obtained by directly imposing the singular value decomposition (SVD) on $\hat{\mathbf{U}} = \frac{1}{\sqrt{m}}[\mathbf{U}^1, \cdots, \mathbf{U}^m] \in \mathbb{R}^{n\times 2mt}$. By doing so, we can leverage the useful information provided by both the consistent and complementary parts to enhance the discriminability of our proposed ESTMC-C^2.

The optimization of Eq. (36) is similar to ESTMC, so we do not repeat it here.

Remark 1 ***[The benefits of CGR]*** *In ESTMC-C^2, the CGR is used to efficiently capture the consistent geometric information. Unlike previous work [36], [60], CGR preserves the geometric information by capturing the geometric structure from the unified anchors instead of the samples, which brings the following benefits: 1) The computational complexity of the adjacency matrix in anchor space* $\mathcal{O}(n)$ *is much lower than that in samples space* $\mathcal{O}(n^2)$*. 2) The CGR captures geometric information in a unified anchor space rather than different views, which helps the dictionary learning methods to learn a consistent clustering structure.*

Remark 2 ***[The benefits of TER and concatenation-fusion strategy]*** *In previous work, they have focused on capturing consistent information among views, but for some special objects (e.g., boundary samples and samples of similar classes), it becomes difficult to achieve accurate classification using consistent information alone. Therefore, our proposed TER could explicitly capture the complementary information across views and the concatenation-fusion strategy can integrate the consistent parts and the complementary parts to improve the discriminative properties of these special objects.*

4 Model Analysis

4.1 Convergence Analysis

The convergence of ESTMC and ESTMC-C^2 is ensured by the following Theorem 3, and according to Theorem 3, the sequence obtained by Algorithm 1 converges to a stationary Karush–Kuhn–Tucker (KKT) critical point [61]. The proof of Theorem 3 is in Appendix.

Theorem 3 *Let* $\{\mathcal{L}_{1(k)} = (\mathbf{Z}_{anc(k)}^v, \mathbf{P}_{(k)}^v, \mathbf{A}_{(k)}, \mathbf{E}_{(k)}^v, \mathbf{Y}_{(k)}^v, \mathcal{W}_{(k)}, \mathcal{G}_{(k)})\}_{k=1}^{\infty}$ *and* $\{\mathcal{L}_{2(k)} = (\mathbf{Z}_{con(k)}^v, \mathbf{Z}_{com(k)}^v, \mathbf{P}_{(k)}^v, \mathbf{A}_{(k)}, \mathbf{E}_{(k)}^v, \mathbf{Y}_{(k)}^v, \mathcal{W}_{(k)}, \mathcal{G}_{(k)})\}_{k=1}^{\infty}$ *be the sequences generated by ESTMC and ESTMC-C^2, then the sequences* $\{\mathcal{L}_{1(k)}\}_{k=1}^{\infty}$ *and* $\{\mathcal{L}_{2(k)}\}_{k=1}^{\infty}$ *meet the following two principles:*

1). $\{\mathcal{L}_{1(k)}\}_{k=1}^{\infty}$ *and* $\{\mathcal{L}_{2(k)}\}_{k=1}^{\infty}$ *are bounded;*
2). *Any accumulation point of* $\{\mathcal{L}_{1(k)}\}_{k=1}^{\infty}$ *and* $\{\mathcal{L}_{2(k)}\}_{k=1}^{\infty}$ *is a KKT point.*

4.2 Complexity Analysis

Time Complexity Analysis: For the proposed ESTMC, the main time complexity is focused on solving for the variables $\mathcal{G}, \mathbf{E}, \mathbf{Z}^v, \mathbf{P}^v$, and $\mathbf{A}$. For tensor $\mathcal{G}$, each update requires FFT, inverse FFT, and t-SVD operations, corresponding to a time complexity of $\mathcal{O}(nmt\log(nm) + nm^2t)$. The update of the variables $\mathbf{Z}^v$ requires $\mathcal{O}(nld^v)$. For $\mathbf{E}$, it needs $\mathcal{O}(nd)$ for each iteration, where $d = \sum_{v=1}^{m} d^v$. And the time complexity of updating $\mathbf{P}^v$ and $\mathbf{A}$ are $\mathcal{O}(nld^v + d^v l^2)$ and $\mathcal{O}(ntmd^v)$, respectively. Hence, considering $t \ll n$, the time complexity of ESTMC is $\mathcal{O}(nmtlog(nm) + nd)$. The optimization of ESTMC-C^2 is similar to ESTMC, so ESTMC-C^2 also enjoys the time complexity $\mathcal{O}(nmtlog(nm) + nd)$. Compared to the traditional tensor-based approaches with a complexity of $\mathcal{O}(n^2log(n))$, ESTMC and ESTMC-C^2 are more efficient and scalable to large-scale datasets.

Storage Complexity Analysis: For our proposed ESTMC and ESTMC-C^2, the major memory costs are various matrices and tensors. From the optimization section, the storage complexity of our proposed methods is $\mathcal{O}((d+mc)n)$, which is linear to the sample size. Compared to existing tensor methods(e.g., TLS$_p$NM-MSC [35] and TBGL-MVC [38]) that have a storage complexity of $O(n^2)$, our proposed methods facilitate the efficient processing of large-scale datasets with reduced resource consumption.

5 Experiment

In this section, we design extensive experiments to demonstrate the effectiveness and superiority of our proposed ESTMC and ESTMC-C^2. Specifically, we compare them with several state-of-the-art algorithms on various datasets with different types and sizes. All the experiments are implemented on a computer with a 2.90GHz i7-10700 CPU and

32GB RAM, Matlab R2020b. The code is publicly available at: https://github.com/jijintian/ESTMC.

TABLE 3: Summary of the benchmark datasets.

Dataset	Sam./Clu.	Dimensions
NGs	500 / 5	2000/2000/2000
BBCSport	544/5	3183/3203
HW	2000 / 10	216/76/64/6/240/47
BDGP	2500 /5	1000/500
CCV	6773 / 20	20/20/20
Caltech101-all	9144 /102	48/40/154/1984/512/928
Aloi-100	11025 / 100	77/13/64/125
Animal	11673 / 20	2689/2000/2001/2000
NUS	30000 / 31	65/226/145/74/129
AWA	30475 / 50	2688/2000/252/2000/2000/2000
CIFAR10	50000 / 10	512/2048/1024
Noisy_MNIST	50000 / 10	784/784

5.1 Experimental Settings

Datasets: To verify the effectiveness of the proposed methods, twelve challenging datasets with different properties are used and summarized in Table 3. All datasets are considered in the ideal scenario, where no samples are missing and the data is arranged in a predefined order. More detailed information about these datasets can be found below.

- **NGs**[1]. It is a subset of the 20 Newsgroup dataset, which consists of 500 newsgroup documents. Each raw document is pre-processed with three different methods (three views) and is classified with one of five labels.
- **BBCSport**[2]. It consists of 544 new sports articles in five topical areas (athletics, cricket, football, rugby, and tennis). Each sample is described by three views, and their dimensions are 2582, 2544, and 2465, respectively.
- **HW**[3]. It is a handwritten digits (0-9) dataset, which comes from the UCI repository and consists of 2000 samples. Each sample is represented by six types of features, whose dimensions are respectively 216, 76, 64, 6, 240, and 47.
- **BDGP**[4]. It contains 2500 Drosophila embryo samples in 5 classes. Each sample has two views with dimensions of 1000 and 500.
- **CCV**[5]. It consists of 6773 samples belonging to 20 semantic categories of YouTube videos. Each sample has three views, and the dimensions are all 20.
- **Caltech101-all**. It is a subset of the challenging object database Caltechall [62] dataset and contains 9144 images belonging to 102 categories. Each sample has six types of features, i.e., 48-dim Gabor, 40-dim Wavelet Moments, 254-dim CENTRIST, 1984-dim HOG, 512-dim GIST and 928-dim LBP.
- **Aloi-100**[6]. It is a broadly adopted small objects dataset, which consists of 11025 images of 100 small objects. Each sample is represented by four types of features: RGB, HSV, Color similarity, and Haralick features.
- **Animal**[7]. It consists of 11673 animal pictures belonging to 20 categories, each sample has four views.
- **NUS** [63]. It is a challenging dataset of 30000 object images with 31 categories. Each sample has 5 different feature views.
- **AWA**[8]. It contains 30475 images of 50 animals and collects six common image features to form a six-feature view of each sample.
- **CIFAR10**[9]: It consists of 50,000 tiny images that can be divided into ten mutually exclusive classes. Similar to [64], we extract its features from three different views.
- **Noisy_MNIST** [38]. It has 50,000 samples with 2 views, where the first view is original data, and the second view is constructed by random choosing within-class images with white Gaussian noise.

Baselines: We compare the following ten state-of-the-art multi-view clustering methods with our proposed methods, including **SC-best** [2], **GMC** (2019 TKDE) [7], **CoMSC** (2021 TNNLS) [65], **OMSC** (2022 KDD) [66], **RCAGL** (2024 TKDE) [67], **SMVSC** (2021 ACM MM) [21], **EOMSC-CA** (2022 AAAI) [22], **t-SVD-MSC** (2018 IJCV) [26], **TLS$_p$NM-MSC** (2023 TPAMI) [35], **TBGL-MVC** (2023 TPAMI) [38], and **ARLRR-TU** (2023 TPAMI) [68]. Their detailed distinguishing features are provided in Table 4.

TABLE 4: Detailed characteristics of baselines. Case #1–#3 indicate whether the corresponding algorithm has the ability to capture the high-order correlations, whether it is accelerated by employing the anchor strategy, and whether it has the ability to explicitly capture the complementary information, respectively.

Baselines	Computational complexity	Case# 1	Case# 2	Case# 3
GMC	$\mathcal{O}(n^2)$	✗	✗	✗
CoMSC	$\mathcal{O}(n)$	✗	✓	✗
OMSC	$\mathcal{O}(n)$	✗	✓	✗
RCAGL	$\mathcal{O}(n)$	✗	✓	✗
SWVSC	$\mathcal{O}(n)$	✗	✓	✗
EOMSC-CA	$\mathcal{O}(n)$	✗	✓	✗
t-SVD-MSC	$\mathcal{O}(n^2logn)$	✓	✗	✗
TLSpNM-MSC	$\mathcal{O}(n^2logn)$	✓	✗	✗
TBGL-MVC	$\mathcal{O}(nlogn)$	✓	✓	✗
ARLRR-TU	$\mathcal{O}(n^2logn)$	✓	✗	✗

Metrics: To measure the clustering performance of our proposed algorithms, five metrics are adopted, including Accuracy (ACC), Normalized Mutual Information (NMI), Purity, F-score, and the Adjusted Rand index (ARI). All the metrics indicate better performance with larger values.

1. https://lig-membres.imag.fr/grimal/data.html
2. http://mlg.ucd.ie/datasets/segment.html
3. http://archive.ics.uci.edu/ml/datasets/Multiple+Features
4. https://www.fruitfly.org/
5. http://www.ee.columbia.edu/ln/dvmm/CCV/
6. http://elki.dbs.ifi.lmu.de/wiki/DataSets/MultiView
7. https://github.com/wangsiwei2010/
8. https://cvml.ista.ac.at/AwA/
9. https://www.cs.toronto.edu/ kriz/cifar.html

TABLE 5: Results (mean(std)%) of our proposed methods and other compared methods on twelve datasets. 'OM' indicates the "out-of-memory error". 'NAN' means the 'not a number error'.

Dataset	Metric	SC_{best}	GMC	CoMSC	OMSC	RCAGL	SMVSC	EOMSC -CA	t-SVD -MSC	TLS_pNM -MSC	TBGL -MVC	ARLRR-TU	ESTMC-ETR	ESTMC-C^2
NGs	ACC	26.06(0.30)	98.20(0.00)	85.20(0.00)	73.80(0.00)	65.60(0.00)	74.20(0.00)	63.00(0.00)	90.00(0.00)	**100(0.00)**	34.20(0.00)	**100(0.00)**	99.60(0.00)	**100(0.00)**
	NMI	1.84(0.15)	93.92(0.00)	64.17(0.00)	59.23(0.00)	55.45(0.00)	52.38(0.00)	46.17(0.00)	76.64(0.00)	**100(0.00)**	16.17(0.00)	**100(0.00)**	98.78(0.00)	**100(0.00)**
	Purity	26.06(0.30)	98.20(0.00)	85.20(0.00)	73.80(0.00)	80.20(0.00)	74.20(0.00)	64.80(0.00)	90.00(0.00)	**100(0.00)**	34.80(0.00)	**100(0.00)**	99.60(0.00)	**100(0.00)**
	F-score	22.21(0.12)	96.43(0.00)	72.86(0.00)	65.68(0.00)	60.28(0.00)	58.31(0.00)	54.24(0.00)	82.03(0.00)	**100(0.00)**	25.40(0.00)	**100(0.00)**	99.20(0.00)	**100(0.00)**
	ARI	0.55(0.14)	95.54(0.00)	66.09(0.00)	56.61(0.00)	49.39(0.00)	47.25(0.00)	41.30(0.00)	77.57(0.00)	**100(0.00)**	11.75(0.00)	**100(0.00)**	99.00(0.00)	**100(0.00)**
BBCSport	ACC	50.31(0.38)	80.70(0.00)	87.32(0.00)	45.22(0.00)	55.33(0.00)	52.39(0.00)	46.32(0.00)	34.74(0.00)	**99.82(0.00)**	54.78(0.00)	**99.82(0.00)**	**99.82(0.00)**	**99.82(0.00)**
	NMI	21.01(0.18)	72.26(0.00)	69.03(0.00)	19.63(0.00)	40.17(0.00)	21.77(0.00)	21.57(0.00)	1.84(0.00)	99.29(0.00)	27.75(0.00)	**99.38(0.00)**	**99.38(0.00)**	**99.38(0.00)**
	Purity	55.28(0.38)	84.38(0.00)	87.32(0.00)	51.10(0.00)	60.48(0.00)	53.68(0.00)	50.37(0.00)	23.61(0.00)	**99.82(0.00)**	54.96(0.00)	**99.82(0.00)**	**99.82(0.00)**	**99.82(0.00)**
	F-score	41.33(0.23)	79.43(0.00)	76.56(0.00)	35.57(0.00)	44.54(0.00)	38.50(0.00)	34.86(0.00)	37.86(0.00)	**99.83(0.00)**	47.32(0.00)	**99.83(0.00)**	**99.83(0.00)**	**99.83(0.00)**
	ARI	15.89(0.47)	72.18(0.00)	69.63(0.00)	13.69(0.00)	27.00(0.00)	17.47(0.00)	15.04(0.00)	0.62(0.00)	99.66(0.00)	18.75(0.00)	**99.77(0.00)**	**99.77(0.00)**	**99.77(0.00)**
HW	ACC	73.85(0.00)	88.20(0.00)	77.45(0.00)	80.80(0.00)	87.75(0.00)	82.05(0.00)	76.00(0.00)	**100(0.00)**	99.95(0.00)	**100(0.00)**	99.90(0.00)	99.80(0.00)	**100(0.00)**
	NMI	69.92(0.00)	89.32(0.00)	74.98(0.00)	77.42(0.00)	80.61(0.00)	78.86(0.00)	77.89(0.00)	**100(0.00)**	99.86(0.00)	**100(0.00)**	99.72(0.00)	99.51(0.00)	**100(0.00)**
	Purity	73.85(0.00)	88.20(0.00)	78.45(0.00)	80.80(0.00)	87.75(0.00)	82.05(0.00)	76.20(0.00)	**100(0.00)**	99.95(0.00)	**100(0.00)**	99.90(0.00)	99.80(0.00)	**100(0.00)**
	F-score	65.27(0.00)	86.53(0.00)	70.08(0.00)	73.42(0.00)	78.08(0.00)	75.25(0.00)	73.42(0.00)	**100(0.00)**	99.90(0.00)	**100(0.00)**	99.79(0.00)	99.60(0.00)	**100(0.00)**
	ARI	61.20(0.00)	84.96(0.00)	66.66(0.00)	70.27(0.00)	75.61(0.00)	72.34(0.00)	69.94(0.00)	**100(0.00)**	99.89(0.00)	**100(0.00)**	99.78(0.00)	99.56(0.00)	**100(0.00)**
BDGP	ACC	32.11(0.03)	13.96(0.00)	58.92(0.00)	63.80(0.00)	56.48(0.00)	65.52(0.00)	43.64(0.00)	48.20(0.06)	96.76(0.00)	NAN	60.68(0.00)	98.56(0.00)	**98.60(0.00)**
	NMI	8.02(0.01)	19.40(0.00)	37.55(0.00)	36.18(0.00)	38.30(0.00)	37.36(0.00)	20.26(0.00)	3.46(0.08)	89.47(0.00)	NAN	44.56(0.00)	95.40(0.00)	**95.63(0.00)**
	Purity	35.09(0.03)	71.08(0.00)	61.28(0.00)	63.80(0.00)	57.92(0.00)	65.52(0.00)	47.08(0.00)	39.83(0.08)	96.76(0.00)	NAN	62.24(0.00)	98.56(0.00)	**98.60(0.00)**
	F-score	28.03(0.01)	20.34(0.00)	49.77(0.00)	45.94(0.00)	45.95(0.00)	47.58(0.00)	32.99(0.00)	40.05(0.09)	93.66(0.00)	NAN	50.79(0.00)	97.14(0.00)	**97.22(0.00)**
	ARI	5.66(0.01)	2.10(0.00)	36.53(0.00)	31.82(0.00)	32.16(0.00)	33.45(0.00)	12.44(0.00)	24.99(0.08)	92.08(0.00)	NAN	37.04(0.00)	96.42(0.00)	**96.52(0.00)**
CCV	ACC	13.52(0.30)	10.63(0.00)	16.92(0.00)	24.02(0.00)	24.42(0.00)	22.06(0.00)	24.32(0.00)	47.87(0.06)	59.99(0.00)	14.6(0.00)	30.65(0.00)	74.58(0.00)	**75.05(0.00)**
	NMI	9.40(0.12)	0.38(0.00)	12.28(0.00)	18.55(0.00)	19.05(0.00)	15.88(0.00)	18.59(0.00)	41.37(0.03)	52.53(0.00)	6.79(0.00)	22.17(0.00)	73.63(0.00)	**72.13(0.00)**
	Purity	17.37(0.19)	10.79(0.00)	19.78(0.00)	26.89(0.00)	26.25(0.00)	25.14(0.00)	26.68(0.00)	49.86(0.07)	61.54(0.00)	16.52(0.00)	33.00(0.00)	75.49(0.00)	**77.37(0.00)**
	F-score	8.51(0.16)	10.86(0.00)	10.41(0.00)	13.84(0.00)	13.70(0.00)	13.14(0.00)	13.99(0.00)	33.47(0.07)	45.41(0.00)	6.2(0.00)	16.73(0.00)	65.97(0.00)	**60.61(0.00)**
	ARI	3.13(0.16)	0.02(0.00)	4.84(0.00)	8.37(0.00)	8.22(0.00)	7.35(0.00)	8.52(0.00)	29.65(0.04)	42.25(0.00)	0.92(0.00)	11.94(0.00)	64.00(0.00)	**58.29(0.00)**
Caltech 101-all	ACC	19.5(0.65)	19.50(0.00)	14.09(0.00)	34.43(0.00)	37.34(0.00)	32.16(0.00)	24.70(0.00)	47.39(1.28)	62.99(0.00)	20(0.00)	62.92(0.00)	58.76(0.00)	**64.90(0.00)**
	NMI	40.44(0.38)	23.79(0.00)	31.83(0.00)	40.90(0.00)	47.84(0.00)	40.57(0.00)	21.09(0.00)	71.26(0.50)	84.80(0.00)	18.76(0.00)	**87.96(0.00)**	83.33(0.00)	85.19(0.00)
	Purity	40.34(0.46)	30.12(0.00)	28.41(0.00)	37.24(0.00)	56.14(0.00)	39.57(0.00)	27.64(0.00)	68.76(0.87)	83.40(0.00)	24.51(0.00)	**86.77(0.00)**	83.03(0.00)	85.14(0.00)
	F-score	15.01(0.76)	4.96(0.00)	8.65(0.00)	25.26(0.00)	25.55(0.00)	30.01(0.00)	14.19(0.00)	32.84(1.21)	49.02(0.00)	3.09(0.00)	46.77(0.00)	44.04(0.00)	**49.38(0.00)**
	ARI	13.47(0.77)	-0.42(0.00)	7.28(0.00)	21.76(0.00)	22.22(0.00)	27.68(0.00)	10.02(0.00)	31.65(1.23)	48.1(0.00)	0.55(0.00)	45.91(0.00)	43.14(0.00)	**48.49(0.00)**
Aloi-100	ACC	64.61(1.20)	65.28(0.00)	69.79(0.00)	35.34(0.00)	41.56(0.00)	34.35(0.00)	22.48(0.00)	71.99(1.44)	83.59(0.00)	66.1(0.00)	84.44(0.00)	83.81(0.00)	**89.45(0.00)**
	NMI	80.03(0.37)	81.27(0.00)	82.52(0.00)	68.39(0.00)	65.71(0.00)	60.99(0.00)	56.19(0.00)	83.92(0.43)	90.89(0.00)	69.25(0.00)	92.25(0.00)	92.97(0.00)	**94.31(0.00)**
	Purity	67.32(0.85)	51.59(0.00)	71.66(0.00)	36.40(0.00)	80.15(0.00)	35.58(0.00)	23.47(0.00)	58.03(1.56)	85.26(0.00)	67.41(0.00)	86.24(0.00)	86.54(0.00)	**90.39(0.00)**
	F-score	55.04(1.26)	57.12(0.00)	58.16(0.00)	19.30(0.00)	8.85(0.00)	21.58(0.00)	8.51(0.00)	62.12(1.11)	75.15(0.00)	6.81(0.00)	77.91(0.00)	80.41(0.00)	**84.91(0.00)**
	ARI	54.52(1.29)	56.65(0.00)	57.7(0.00)	17.90(0.00)	7.13(0.00)	20.39(0.00)	6.76(0.00)	61.71(1.13)	77.5(0.00)	10.74(0.00)	79.76(0.00)	80.21(0.00)	**84.76(0.00)**
Animal	ACC	14.44(0.17)	9.99(0.00)	13.08(0.00)	19.07(0.00)	16.03(0.00)	19.49(0.00)	18.87(0.00)	16.77(0.10)	25.28(0.00)	9.16(0.00)	71.90(0.00)	92.25(0.00)	**92.28(0.00)**
	NMI	10.48(0.15)	1.17(0.00)	8.81(0.00)	15.86(0.00)	13.57(0.00)	15.72(0.00)	14.80(0.00)	12.56(0.06)	19.14(0.00)	0.17(0.00)	62.94(0.00)	91.62(0.00)	**91.65(0.00)**
	Purity	17.82(0.15)	10.09(0.00)	16.57(0.00)	21.82(0.00)	23.05(0.00)	21.34(0.00)	21.37(0.00)	20.03(0.07)	27.29(0.00)	9.21(0.00)	72.99(0.00)	93.40(0.00)	**93.43(0.00)**
	F-score	9.14(0.15)	11.07(0.00)	8.29(0.00)	14.17(0.00)	12.19(0.00)	14.89(0.00)	13.82(0.00)	10.40(0.05)	14.04(0.00)	5.86(0.00)	59.58(0.00)	89.09(0.00)	**89.14(0.00)**
	ARI	3.19(0.07)	0.04(0.00)	2.95(0.00)	8.33(0.00)	6.51(0.00)	8.35(0.00)	6.99(0.00)	4.56(0.04)	9.86(0.00)	0(0.00)	57.21(0.00)	88.44(0.00)	**88.50(0.00)**
NUS	ACC	11.50(0.31)	OM	OM	19.93(0.00)	19.30(0.00)	18.44(0.00)	19.51(0.00)	OM	OM	OM	OM	15.44(0.00)	**21.07(0.00)**
	NMI	10.73(0.15)	OM	OM	12.78(0.00)	12.61(0.00)	11.77(0.00)	12.68(0.00)	OM	OM	OM	OM	13.09(0.00)	**20.17(0.00)**
	Purity	24.22(0.28)	OM	OM	23.46(0.00)	**36.18(0.00)**	22.97(0.00)	23.94(0.00)	OM	OM	OM	OM	24.67(0.00)	30.73(0.00)
	F-score	8.27(0.20)	OM	OM	13.25(0.00)	13.49(0.00)	12.30(0.00)	**13.76(0.00)**	OM	OM	OM	OM	9.45(0.00)	12.33(0.00)
	ARI	4.10(0.17)	OM	OM	6.27(0.00)	5.86(0.00)	6.02(0.00)	6.92(0.00)	OM	OM	OM	OM	5.36(0.00)	**8.38(0.00)**
AWA	ACC	8.81(0.15)	OM	OM	9.76(0.00)	9.63(0.00)	9.30(0.00)	8.70(0.00)	OM	OM	OM	OM	92.4(0.00)	**93.82(0.00)**
	NMI	9.38(0.05)	OM	OM	11.55(0.00)	12.03(0.00)	10.63(0.00)	8.75(0.00)	OM	OM	OM	OM	94.89(0.00)	**95.87(0.00)**
	Purity	10.85(0.11)	OM	OM	10.54(0.00)	15.93(0.00)	10.18(0.00)	9.40(0.00)	OM	OM	OM	OM	95.85(0.00)	**96.29(0.00)**
	F-score	4.39(0.05)	OM	OM	6.27(0.00)	5.30(0.00)	6.12(0.00)	5.99(0.00)	OM	OM	OM	OM	91.01(0.00)	**93.36(0.00)**
	ARI	2.08(0.03)	OM	OM	2.73(0.00)	2.66(0.00)	2.83(0.00)	2.53(0.00)	OM	OM	OM	OM	90.8(0.00)	**93.20(0.00)**
CIFAR10	ACC	90.70(0.01)	OM	OM	98.99(0.00)	98.98(0.00)	98.85(0.00)	99.08(0.00)	OM	OM	OM	OM	**100(0.00)**	**100(0.00)**
	NMI	80.53(0.01)	OM	OM	97.31(0.00)	97.29(0.00)	97.03(0.00)	97.48(0.00)	OM	OM	OM	OM	**100(0.00)**	**100(0.00)**
	Purity	90.70(0.01)	OM	OM	98.99(0.00)	98.98(0.00)	98.85(0.00)	99.08(0.00)	OM	OM	OM	OM	**100(0.00)**	**100(0.00)**
	F-score	82.68(0.01)	OM	OM	98.01(0.00)	98.00(0.00)	97.75(0.00)	98.18(0.00)	OM	OM	OM	OM	**100(0.00)**	**100(0.00)**
	ARI	80.75(0.01)	OM	OM	97.79(0.00)	97.78(0.00)	97.50(0.00)	97.97(0.00)	OM	OM	OM	OM	**100(0.00)**	**100(0.00)**
Noisy_ MNIST	ACC	62.82(2.97)	OM	OM	59.49(0.00)	52.00(0.00)	60.28(0.00)	58.65(0.00)	OM	OM	OM	OM	98.78(0.00)	**99.56(0.00)**
	NMI	62.80(1.69)	OM	OM	49.35(0.00)	46.39(0.00)	51.89(0.00)	49.08(0.00)	OM	OM	OM	OM	97.18(0.00)	**98.81(0.00)**
	Purity	69.10(1.31)	OM	OM	61.01(0.00)	56.00(0.00)	60.95(0.00)	58.83(0.00)	OM	OM	OM	OM	98.78(0.00)	**99.56(0.00)**
	F-score	56.66(2.29)	OM	OM	47.78(0.00)	44.10(0.00)	49.24(0.00)	46.72(0.00)	OM	OM	OM	OM	97.59(0.00)	**99.13(0.00)**
	ARI	51.58(2.69)	OM	OM	41.66(0.00)	37.68(0.00)	43.14(0.00)	40.44(0.00)	OM	OM	OM	OM	97.32(0.00)	**99.04(0.00)**

TABLE 6: Average running time (sec.) comparison of multi-view clustering methods under datasets with more than 6,000 samples.' OM' indicates the "out-of-memory error".

Dataset	GMC	CoMSC	OMSC	RCAGL	SMVSC	EOMSC -CA	t-SVD -MSC	TLS_pNM -MSC	TBGL -MVC	ARLRR-TU	ESTMC-ETR	ESTMC-C^2
CCV	476.03	95.58	42.73	15.96	105.01	13.51	2630.27	2459.58	2600.69	2438.34	**4.17**	4.81
Caltech101-all	461.73	1549.456	2961.70	92.07	980.71	**53.56**	13840.56	31219.72	18259.95	17385.06	125.37	107.5
Aloi-100	633.14	748.9917	914.05	56.80	658.18	86.38	14494.13	64684.79	42223.69	12010.17	**41.85**	42.36
Animal	676.36	483.443	134.07	55.71	136.19	**29.63**	16332.87	35213.18	39613.85	21256.57	198.40	150.39
NUS	OM	OM	432.21	355.34	429.65	106.52	OM	OM	OM	OM	**66.02**	102.85
AWA	OM	OM	4044.14	288.13	1985.10	**71.41**	OM	OM	OM	OM	633.55	730.72
CIFAR10	OM	OM	490.66	117.53	547.36	**97.29**	OM	OM	OM	OM	286.73	250.07
Noisy_MNIST	OM	OM	496.10	109.35	528.41	**99.57**	OM	OM	OM	OM	130.19	118.41

Parameter Setting: For our proposed methods, three trade-off parameters need to be tuned, we set the value ranges of λ, β, α to $[10^{-6}, 10^{-5}, \cdots, 1]$. For the anchor space, the dimension l and anchor number t are both set to the number of classes c. Due to the varying parameter ranges of the different non-convex approximation functions, our proposed ESTMC and ESTMC-C^2 adopt ETR with $\delta = \{10^{-2}, 10^{-1}, 1\}$, more non-convex surrogate functions will be discussed in the ablation part. For all the baselines, their related parameters are tuned with the ranges suggested in the original paper and the best results are shown. Since most algorithmic codes employ fixed random numbers, all methods in the experiment are run 10 times and the average is reported.

5.2 Clustering Results

The results of proposed ESTMC and ESTMC-C^2 and all baselines are shown in Table 5 and Table 6, where the **best** and second best values in all algorithms are denoted by bold values and underlined values, respectively. ESTMC-ETR is ESTMC using the proposed tensor rank ETR. From all the results, we have the following observations.

From a global perspective, ESTMC-ETR and ESTMC-C^2 show commendable performance on all datasets. More importantly, ESTMC-ETR stands out with optimal results on the CIFAR10 dataset, while ESTMC-C^2 obtains the ideal clustering results on three distinct datasets (*i.e.*, NGs, HW, and CIFAR10). Notably, our proposed methods significantly enhance performance on challenging datasets, such as AWA, where they both obtain over an 80% improvement in all five metrics (ACC, NMI, Purity, F-score, and ARI) compared to competing algorithms. On the Animal dataset, ESTMC-C^2 reaches the best performance and gains significant improvements over ARLRR-TU over five metrics, *i.e.*, 20.38%, 28.71%, 20.44%, 29.56%, and 31.29%. Furthermore, our proposed algorithms can effectively handle these large-scale datasets (AWA, NUS, CIFAR10, and Noisy_MNIST) with limited resources compared to existing tensor algorithms.

Then, in our detailed analysis of Table 5, we observed the following observations:

1) The single-view spectral clustering (SC-best) is inferior to multi-view clustering algorithms in most cases. The reason for this phenomenon may be that multi-view clustering methods can exploit consistent and complementary information among the views to obtain a clear clustering structure. In contrast, single-view spectral clustering relies solely on one view, limiting its performance.

2) Tensor-based approaches, such as TLS$_p$NM-MSC, ARLRR-TU, and ESTMC-C^2, demonstrate better performances than matrix-oriented approaches on most datasets. This is attributed to the ability of tensor-based methods to capture high-order correlation, enhancing the learning of consistent clustering structures across views, which matrix-oriented methods cannot achieve.

3) Although our proposed ESTMC framework and TBGL-MVC are all anchor-based methods, ESTMC shows consistently competitive results on different datasets. For example, on NGs, BBCSport, and HW datasets, the clustering stability of TBGL-MVC is inferior to ESTMC. This is mainly because ESTMC combines anchor learning and low-rank representation learning into a unified framework to learn the optimal anchors and the optimal representations simultaneously, thus the sub-optimal solution problem caused by the pre-selected anchor method is avoided.

4) The proposed ESTMC-C^2 presents superior performances than other tensor-based methods on all datasets. Especially compared to ESTMC, which uses a similar framework, the clustering results on most of the datasets are improved. For example, on the Caltech101-all dataset, ESTMC-C^2 gains significant improvements around 6.14%, 1.86%, 2.11%, 5.34%, 5.35% of five metrics, respectively. This is because complementary information captured by ESTMC-C^2 can significantly improve the discriminability of some special samples like boundary samples.

Based on the runtime results in Table 6, additional insights include:

1) Compared to other tensor-based methods (t-SVD-MVC, TLS$_p$NM-MSC), our proposed ESTMC and ESTMC-C^2 both have the ability to efficiently handle large-scale datasets. This is because the anchor representation strategy can reduce the computational complexity of tensor-related operations from $\mathcal{O}(n^2)$ to $\mathcal{O}(nlog(n))$. More importantly, the storage complexity of proposed ESTMC and ESTMC-C^2 is $\mathcal{O}(n)$, which is much lower than $\mathcal{O}(n^2)$ in tensor-based methods.

2) Although TBGL-MVC is also an anchor-based method, the number of anchors is much higher than that in our ESTMC and ESTMC-C^2 and the storage complexity reaches $\mathcal{O}(n^2)$, thus limiting its performance and efficiency on large-scale datasets.

3) Despite ESTMC and ESTMC-C^2 taking a little longer time as compared with the anchor-based and matrix-oriented methods, e.g., EOMSC-CA, OMSC, RCAGL, and SMVSC, they show excellent performance that significantly beats these methods on all challenging datasets. This shows that the combination of anchor strategy and low-rank tensor learning strategy can achieve both efficiency and accuracy.

5.3 Ablation Study

Influence of different non-convex surrogate functions: To investigate the impact of various non-convex functions, we instantiate the proposed GNTR in ESTMC with four popular forms (*e.g.*, ℓ_δ-norm, Geman, Laplace, and our proposed ETR), which are named ESTMC-ℓ_δ, ESTMC-Gem, ESTMC-Lap and ESTMC-ETR, respectively. Their parameter ranges are all set to $[10^{-3}, 10^{-2}, \cdots, 10]$. Meanwhile, in order to demonstrate the difference between the non-convex function and the traditional TNN, the TNN-based variant ESTMC-TNN is also constructed. We compare the clustering performance of all variants on NGs, BBCSport, HW, BDGP, CCV, Caltech101-all, and Aloi-100 datasets. The clustering results are given in Table 7. It is observed that the methods using non-convex functions (*i.e.,* ESTMC-ETR, ESTMC-ℓ_δ, ESTMC-Gem and ESTMC-Lap) outperform ESTMC-TNN in most cases, the reason for this is that low-rank constraints based on non-convex approximation functions provide appropriate penalty strengths for singular values of different sizes, thus ensuring that the main structure of the data is not corrupted. ESTMC-ETR, ESTMC-ℓ_δ, ESTMC-Gem, and ESTMC-Lap show satisfactory performance on these

TABLE 7: Clustering performance with different surrogate functions of rank.

Variants	NGs			BBCSport			HW			BDGP			Aloi-100		
	ACC	MNI	Purity	ACC	MNI	Purity	ACC	MNI	Purity	ACC	MNI	Purity	ACC	MNI	Purity
ESTMC-TNN	98.80	96.39	98.80	96.14	88.43	96.14	99.45	98.64	99.45	97.76	93.46	97.76	78.33	90.21	80.06
ESTMC-ℓ_δ	99.40	97.91	99.40	99.26	97.43	99.26	99.70	99.18	99.70	**98.60**	**95.51**	**98.60**	82.36	92.43	85.14
ESTMC-Gem	99.20	97.21	99.20	98.16	93.78	98.16	99.20	98.17	99.20	98.56	95.40	98.56	**84.69**	**93.11**	**87.00**
ESTMC-Lap	99.40	97.91	99.40	99.26	97.23	99.26	99.70	99.18	99.70	97.64	93.37	97.64	83.31	92.75	85.97
ESTMC-ETR	**99.60**	**98.78**	**99.60**	**99.82**	**99.38**	**99.82**	**99.80**	**99.51**	**99.80**	98.56	95.40	98.56	83.81	92.97	86.54

datasets, the clustering performances (ACC, NMI, Purity) differ within 2%. From the details, different non-convex approximation functions achieve the best clustering performance on specific datasets, for example, our proposed ETR on the NGs, BBCSport, and HW datasets, and ESTMC-Gem based on the Geman function on the Aloi-100 dataset. This phenomenon suggests that low-rank constraints based on different non-convex approximation functions can well capture the low-rank structure of multi-view data and that different functions have their different application areas.

TABLE 8: Ablation study on eight datasets (ACC(%)).

Dataset	w/o TER	w/o CGR	w/o CS	ESTMC-C^2
NGs	99.80	**100**	93.80	**100**
BBCSport	98.90	**99.82**	**99.82**	**99.82**
HW	97.60	99.50	99.70	**100**
BDGP	98.32	98.50	98.52	**98.6**
CCV	72.30	71.90	65.20	**75.05**
Caltech101-all	59.10	49.22	61.83	**64.9**
Aloi-100	81.06	85.19	86.61	**89.45**
Animal	88.97	92.26	89.19	**92.28**

Influence of TER, CGR, and Concatenation-fusion Strategy: The ESTMC-C^2 model has three important components: Tensorial Exclusive Regularization (TER), Consistent Geometric Regularization (CGR) and the Concatenation-fusion Strategy (CS). To demonstrate the effectiveness of ESTMC-C^2, we conduct the comparison experiments on eight datasets (NGs, BBCSport, HW, BDGP, CCV, Caltech101-all, Aloi-100, and Animal), systematically removing each component while retaining the other, resulting in variants labeled as w/o TER, w/o CGR, and w/o CS. As depicted in Table 8, ESTMC-C^2 consistently outperforms those variants. This is due to its ability to capture both complementary and consistent information across views, making it adaptable to diverse datasets with varying distributions. The performance of those variants varies across datasets, For example, they achieve satisfactory results on BBCSport and BDGP datasets, but their performance is less effective on the Caltech101-all dataset. This discrepancy can be attributed to the roles of CGR and TER in capturing nonlinear structures and view-specific information. For datasets with clear sample distributions, the influence of these regularization terms is negligible. However, they demonstrate significant utility in improving clustering performance in more complex datasets, where chaotic sample distributions are present, by offering comprehensive and structured information.

5.4 Model Analysis

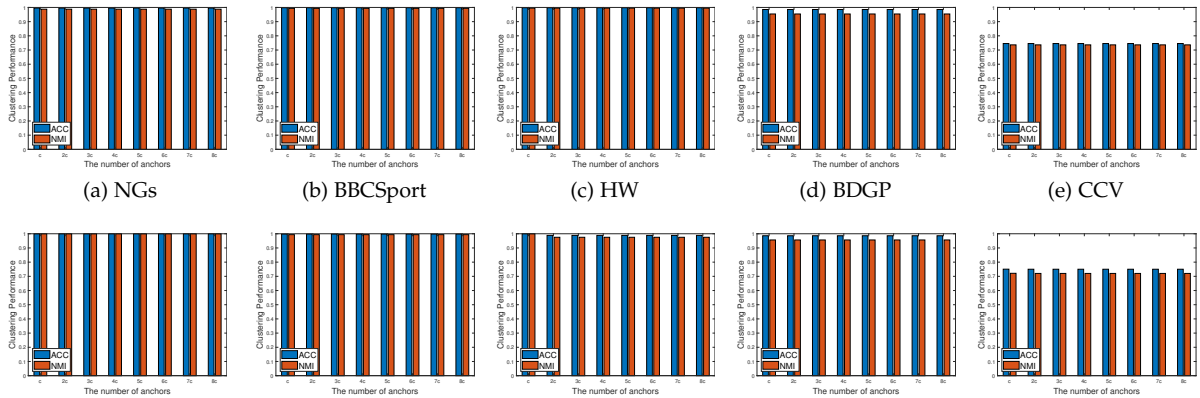
Effect of the Number of Anchors: We empirically analyze the influence of the number of anchors for clustering on five datasets (NGs, BBCSport, HW, BDGP, and CCV). Specifically, we vary the number of anchors within the range $\{c, 2c, 3c, \cdots, 8c\}$, where c is the number of clusters, and the clustering performance (ACC and NMI) is shown in Fig. 5. The results indicate that the choice of anchor numbers has a minimal impact on the clustering performance of our proposed models. Both ESTMC and ESTMC-C^2 consistently achieve satisfactory performance with different values of parameter t. This is because when $t \geq c$, the anchors learned by our models can effectively cover the entire sample space. Consequently, in practical experiments, fixing the number of anchors at c can alleviate the burden of parameter tuning during training.

Parameters Analysis: In our proposed ESTMC-C^2, there are three balance parameters λ, β and α. we empirically employ the grid search strategy to determine the optimal parameters in each dataset, where the search range for all parameters is $\{10^{-6}, 10^{-5}, \cdots, 1\}$. Fig. 6 shows the effect of varying parameters on clustering performance (ACC and NMI) over five datasets (NGs, BBCSport, HW, BDGP, and CCV). We can observe that the performance of the proposed ESTMC-C^2 is relatively insensitive to these three parameters in certain ranges. Since determining the optimal parameters remains an open challenge for unsupervised clustering, it is not necessary to determine the exact value of each parameter, but rather to select a small range of parameters. So the search range for these three parameters can be narrowed down, where the range of λ, α and β can be set to $\{10^{-4}, 10^{-3}, \cdots, 10^{-1}\}$, $\{10^{-2}, 10^{-1}, 1\}$, and $\{10^{-6}, 10^{-5}, \cdots, 10^{-3}\}$, respectively.

Convergence Analysis: The convergence of ESTMC and ESTMC-C^2 has been proved in the previous section and is guaranteed by the Theorem 3. In this section, we use the stopping criteria to further verify the stability of our algorithms and the stopping criteria used here are Reconstruction Error (RE) and Match Error (ME) [26]: $RE = \max_v \|\mathbf{X}^v - \mathbf{X}^v\mathbf{Z}^v - \mathbf{E}^v\|_\infty$ and $ME = \|\mathcal{Z} - \mathcal{G}\|_\infty$, respectively. We show the variation curves of these two stopping criteria on five datasets, which are shown in Fig. 7. The values of RE and ME drop rapidly to 0 and remain stable on all datasets, and the number of iterations on these datasets is within 25 steps, indicating the excellent convergence of our algorithms.

6 Conclusion

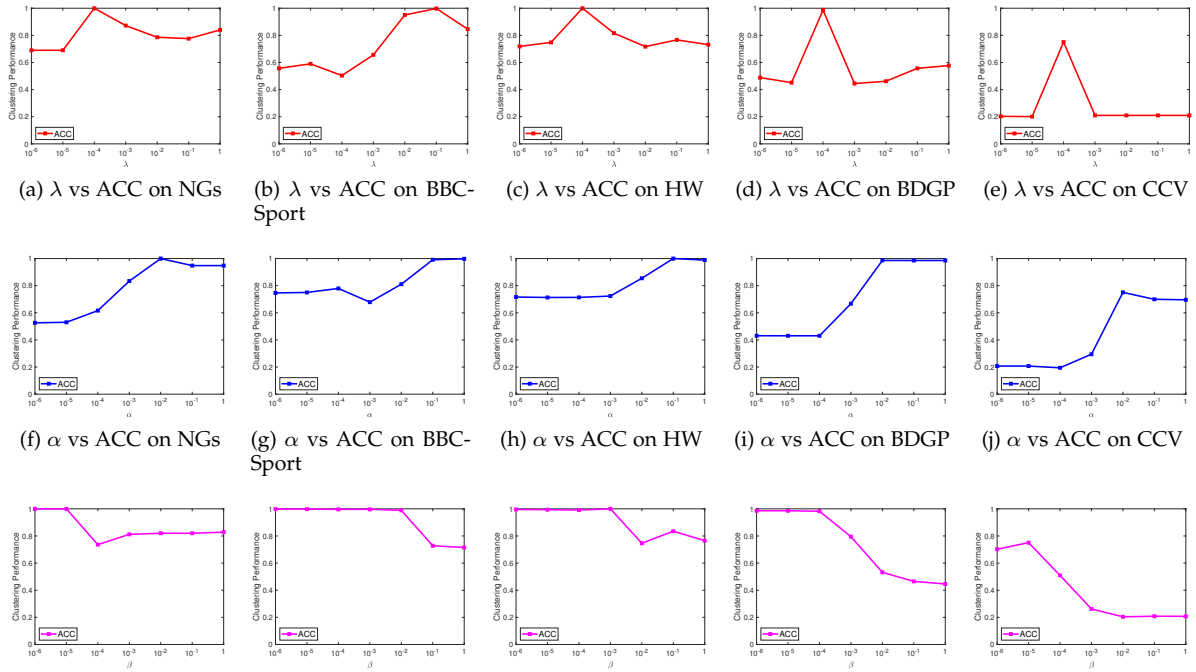
In this paper, we first propose an efficient and scalable framework (ESTMC) for subspace-based TMC tasks.



(a) NGs (b) BBCSport (c) HW (d) BDGP (e) CCV

(f) NGs (g) BBCSport (h) HW (i) BDGP (j) CCV

Fig. 5: The performance (i.e., ACC and NMI) of our proposed methods on five datasets by varying the number of anchors. (a)-(e) are variation curves for ESTMC, (f) - (j) for ESTMC-C^2.

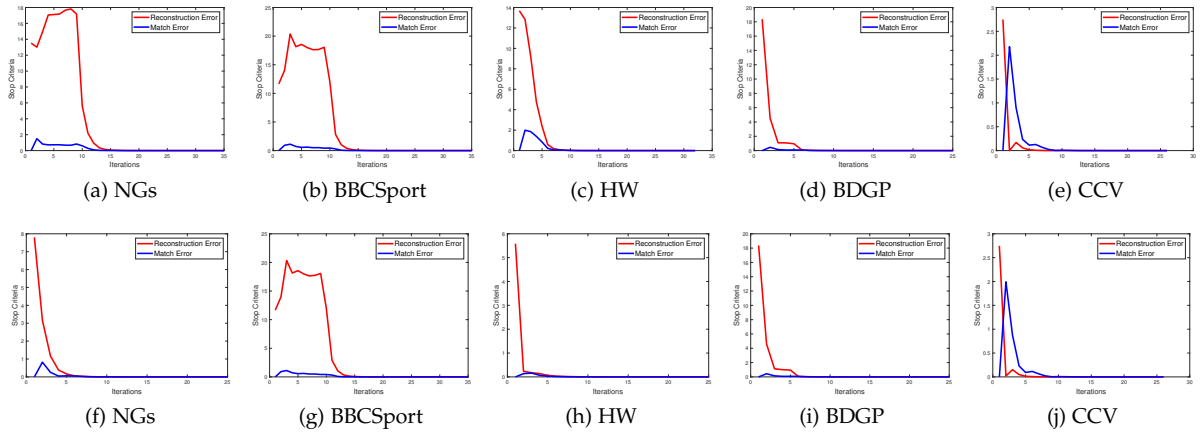


(a) λ vs ACC on NGs (b) λ vs ACC on BBC-Sport (c) λ vs ACC on HW (d) λ vs ACC on BDGP (e) λ vs ACC on CCV

(f) α vs ACC on NGs (g) α vs ACC on BBC-Sport (h) α vs ACC on HW (i) α vs ACC on BDGP (j) α vs ACC on CCV

(k) β vs ACC on NGs (l) β vs ACC on BBC-Sport (m) β vs ACC on HW (n) β vs ACC on BDGP (o) β vs ACC on CCV

Fig. 6: Parameter sensitivity analysis of ESTMC-C^2 on five datasets in terms of ACC. (a)-(e) λ, (f)-(j) α, (k)-(o) β.



(a) NGs (b) BBCSport (c) HW (d) BDGP (e) CCV

(f) NGs (g) BBCSport (h) HW (i) BDGP (j) CCV

Fig. 7: Convergence Analysis: The stop criteria (i.e., RE and ME) variation curves on five datasets. (a)-(e) are convergence curves for ESTMC, (f) - (j) for ESTMC-C^2.

ESTMC integrates anchor representation learning, low-rank tensor learning, and a Generalized Non-convex Tensor Rank (GNTR) into a unified framework, advancing the capabilities of traditional TMC methods on large-scale datasets. To better exploit the discriminative information, we have developed a specialized model, ESTMC-C^2, which employs three distinct regularization terms (ETR, CGR, and TER) to explicitly capture both consistency and complementarity across views. Efficient algorithms are designed to solve our objectives, which have both resource-saving complexity and strict convergence properties. Extensive experiments on twelve challenging datasets demonstrate the effectiveness of our proposed models.

APPENDIX

PROOF OF THEOREM 3

Since the proof procedures for $\{\mathcal{L}_{1(k)}\}_{k=1}^{\infty}$ and $\{\mathcal{L}_{2(k)}\}_{k=1}^{\infty}$ are similar, we just present the detailed proof of $\{\mathcal{L}_{1(k)}\}_{k=1}^{\infty}$, we first introduce two important lemmas.

Lemma 2 *[69] Let $\mathcal{H}$ be a real Hilbert space endowed with an inner product $\langle\cdot,\cdot\rangle$, a norm $\|\cdot\|$ with the dual norm $\|\cdot\|^{dual}$, and $y \in \partial\|x\|$, where $\partial f(\cdot)$ is the subgradient of $f(\cdot)$. Then we have $\|y\|^{dual} = 1$ if $x \neq 0$, and $\|y\|^{dual} \leq 1$ if $x = 0$.*

Lemma 3 *[70] Suppose that $F : \mathbb{R}^{m\times n} \to \mathbb{R}$ is defined as $F(\mathbf{X}) = f \circ \boldsymbol{\sigma}(\mathbf{X}) = f(\sigma_1(\mathbf{X}),\ldots,\sigma_r(\mathbf{X}))$, where $\mathbf{X} = \mathbf{U}Diag(\boldsymbol{\sigma}(\mathbf{X}))\mathbf{V}^T$ is SVD of matrix $X \in \mathbb{R}^{m\times n}, r = \min(m,n)$, and $f(\cdot) : \mathbb{R}^r \to \mathbb{R}$ be differentiable and absolutely symmetric at $\boldsymbol{\sigma}(\mathbf{X})$. Then, the subdifferential of $F(\mathbf{X})$ at $\mathbf{X}$ is*

$$\frac{\partial F(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{U}Diag(\partial f(\boldsymbol{\sigma}(\mathbf{X})))\mathbf{V}^T, \tag{37}$$

where $\partial f(\boldsymbol{\sigma}(\mathbf{X})) = (\frac{\partial f(\sigma_1(\mathbf{x}))}{\partial \mathbf{X}}),\ldots,\frac{\partial f(\sigma_r(\mathbf{x}))}{\partial \mathbf{X}})$.

Proof **1).** Proof of $1st$ part: On the $k+1$ iteration, from the updating rule of $\mathbf{E}_{(k+1)}$, the first-order optimal condition should be satisfied.

$$0 \in \lambda\partial\left\|\mathbf{E}_{(k+1)}\right\|_{2,1} + \mu_{(k)}(\mathbf{E}_{(k+1)} - \hat{\mathbf{E}}), \tag{38}$$

where $\mathbf{E}_{(k+1)} = [\mathbf{E}^1_{(k+1)},\cdots,\mathbf{E}^m_{(k+1)}]^T$. With the updating rule of $\mathbf{Y}^v_{(k+1)}$:

$$\mathbf{Y}^v_{(k+1)} = \mathbf{Y}^v_{(k)} + \mu(\mathbf{X}^v - \mathbf{Z}^v_{anc(k+1)}\mathbf{A}_{(k+1)}\mathbf{P}^v_{(k+1)} - \mathbf{E}^v_{(k+1)}), \tag{39}$$

we obtain $\mu_{(k)}(\hat{\mathbf{E}} - \mathbf{E}_{(k+1)}) = [\mathbf{Y}^1_{(k+1)},\cdots,\mathbf{Y}^m_{(k+1)}]^T$.

Then, from the definition of $\ell_{2,1}$ norm, the following equation can be derived,

$$\frac{1}{\lambda}[\mathbf{Y}_{(k+1)}]_{:,j} \in \partial\left\|[\mathbf{E}_{(k+1)}]_{:,j}\right\|_2, \tag{40}$$

where $[\mathbf{Y}_{(k+1)}]_{:,j}$ and $[\mathbf{E}_{(k+1)}]_{:,j}$ are the j-th columns of $\mathbf{Y}_{(k+1)}$ and $\mathbf{E}_{(k+1)}$. And the ℓ_2 norm is self-dual, so based on the Lemma 2, $\left\|\frac{1}{\lambda}[\mathbf{Y}_{(k+1)}]_{:,j}\right\|_2 \leq 1$, which means that the sequence $\{\mathbf{Y}^v_{(k+1)}\}_{k=1}^{\infty}$ is bounded.

Then, to prove the sequence $\{\mathcal{W}_{(k+1)}\}_{k=1}^{\infty}$ is bounded. We obtain the first-order optimality condition of $\mathcal{G}$ in the updating rule,

$$\mathcal{W}_{(k+1)} \in \partial\left\|\mathcal{G}_{(k+1)}\right\|_{GNTR} \tag{41}$$

Let $\mathcal{U} * \mathcal{S} * \mathcal{V}^T$ be the t-SVD of tensor $\mathcal{G}$. According to the Lemma 3 and Definition 3,

$$\begin{aligned}
&\left\|\partial\left\|\mathcal{G}_{(k+1)}\right\|_{GNTR}\right\|_F^2 \\
&= \left\|\frac{1}{n_3}\mathcal{U} * \text{ifft}(\partial\psi(\mathcal{S}_f,\delta),[],3) * \mathcal{V}^T\right\|_F^2 \\
&= \frac{1}{n_3^3}\left\|\partial\psi(\mathcal{S}_f,\delta)\right\|_F^2 \\
&\leq \frac{1}{n_3^3}\sum_{i=1}^{n_3}\sum_{j=1}^{min(n_1,n_2)}[\partial\psi(\mathcal{S}^v_f(j,j),\delta)]^2.
\end{aligned} \tag{42}$$

Since the singular values of a finite matrix are bounded, $\partial\psi(\mathcal{S}^v_f(j,j),\delta)$ is also bounded, which implies $\partial\left\|\mathcal{G}_{(k+1)}\right\|_{GNTR}$ is bounded. From Eq. (41), we can obtain the sequence $\{\mathcal{W}_{(k+1)}\}_{k=1}^{\infty}$ is also bounded.

Moreover, based on the iterative solution of the proposed Algorithm 1, we can infer that

$$\begin{aligned}
&\mathcal{L}_{\mu_{(k)},\rho_{(k)}}(\mathbf{Z}^v_{anc(k+1)},\mathbf{A}_{(k+1)},\mathbf{P}^v_{(k+1)},\mathbf{E}_{(k+1)},\mathcal{G}_{(k+1)},\mathbf{Y}^v_{(k)},\mathcal{W}_{(k)}) \\
&\leq \mathcal{L}_{\mu_{(k)},\rho_{(k)}}(\mathbf{Z}^v_{anc(k)},\mathbf{A}_{(k)},\mathbf{P}^v_{(k)},\mathbf{E}_{(k)},\mathcal{G}_{(k)},\mathbf{Y}^v_{(k)},\mathcal{W}_{(k)}) \\
&= \mathcal{L}_{\mu_{(k-1)},\rho_{(k-1)}}(\mathbf{Z}^v_{anc(k)},\mathbf{A}_{(k)},\mathbf{P}^v_{(k)},\mathbf{E}_{(k)},\mathcal{G}_{(k)},\mathbf{Y}^v_{(k-1)},\mathcal{W}_{(k-1)}) \\
&+ \frac{\rho_{(k)}+\rho_{(k-1)}}{2\rho^2_{(k-1)}}\left\|\mathcal{W}_{(k)} - \mathcal{W}_{(k-1)}\right\|_F^2 \\
&+ \frac{\mu_{(k)}+\mu_{(k-1)}}{2\mu^2_{(k-1)}}\sum_{v=1}^{m}\left\|\mathbf{Y}^v_{(k)} - \mathbf{Y}^v_{(k-1)}\right\|_F^2,
\end{aligned} \tag{43}$$

Thus, summing right side of (43) form $k=1$ to n,

$$\begin{aligned}
&\mathcal{L}_{\mu_{(k)},\rho_{(k)}}(\mathbf{Z}^v_{anc(k+1)},\mathbf{A}_{(k+1)},\mathbf{P}^v_{(k+1)},\mathbf{E}_{(k+1)},\mathcal{G}_{(k+1)},\mathbf{Y}^v_{(k)},\mathcal{W}_{(k)}) \\
&\leq \mathcal{L}_{\mu_0,\rho_0}(\mathbf{Z}^v_1,\mathbf{A}_1,\mathbf{P}^v_1,\mathbf{E}_1,\mathcal{G}_1,\mathbf{Y}^v_0,\mathcal{W}_0) \\
&+ \sum_{k=1}^{n}\frac{\rho_{(k)}+\rho_{(k-1)}}{2\rho^2_{(k-1)}}\left\|\mathcal{W}_{(k)} - \mathcal{W}_{(k-1)}\right\|_F^2 \\
&+ \sum_{k=1}^{n}\left(\frac{\mu_{(k)}+\mu_{(k-1)}}{2\mu^2_{(k-1)}}\sum_{v=1}^{m}\left\|\mathbf{Y}^v_{(k)} - \mathbf{Y}^v_{(k-1)}\right\|_F^2\right)
\end{aligned} \tag{44}$$

Observe that

$$\sum_{k=1}^{n}\frac{\mu_{(k)}+\mu_{(k+1)}}{2\mu^2_{(k-1)}} < \infty, \quad \sum_{k=1}^{n}\frac{\rho_{(k)}+\rho_{(k+1)}}{2\rho^2_{(k-1)}} < \infty \tag{45}$$

and $\mathcal{L}_{\mu_0,\rho_0}(\mathbf{Z}^v_1,\mathbf{A}_1,\mathbf{P}^v_1,\mathbf{E}_1,\mathcal{G}_1,\mathbf{Y}^v_0,\mathcal{W}_0)$ is finite, and the sequences $\{\mathbf{Y}^v_{(k)}\}_{k=1}^{\infty}, \{\mathcal{W}_{(k)}\}_{k=1}^{\infty}$, $\sum_{k=1}^{n}\frac{\mu_{(k)}+\mu_{(k+1)}}{2\mu^2_{(k-1)}}$ and $\sum_{k=1}^{n}\frac{\rho_{(k)}+\rho_{(k+1)}}{2\rho^2_{(k-1)}}$ are all bounded. So we can obtain $\mathcal{L}_{\mu_{(k)},\rho_{(k)}}(\mathbf{Z}^v_{anc(k+1)},\mathbf{A}_{(k+1)},\mathbf{P}^v_{(k+1)},\mathbf{E}_{(k+1)},\mathcal{G}_{(k+1)},\mathbf{Y}^v_{(k)},\mathcal{W}_{(k)})$ is bounded.

Note that

$$\begin{aligned}
&\mathcal{L}_{\mu_{(k)},\rho_{(k)}}(\mathbf{Z}^v_{anc(k+1)},\mathbf{A}_{(k+1)},\mathbf{P}^v_{(k+1)},\mathbf{E}_{(k+1)},\mathcal{G}_{(k+1)},\mathbf{Y}^v_{(k)},\mathcal{W}_{(k)}) \\
&= \left\|\mathcal{G}_{(k+1)}\right\|_{GNTR} + \lambda\left\|\mathbf{E}_{(k+1)}\right\|_{2,1} \\
&+ \sum_{v=1}^{m}(\langle\mathbf{Y}^v_{(k)},\mathbf{X}^v - \mathbf{Z}^v_{anc(k+1)}\mathbf{A}_{(k+1)}\mathbf{P}^v_{(k+1)} - \mathbf{E}^v_{(k+1)}\rangle \\
&+ \frac{\mu_{(k)}}{2}\left\|\mathbf{X}^v - \mathbf{Z}^v_{anc(k+1)}\mathbf{A}_{(k+1)}\mathbf{P}^v_{(k+1)} - \mathbf{E}^v_{(k+1)}\right\|_F^2) \\
&+ \langle\mathcal{W}_{(k)},\mathcal{Z}_{(k+1)} - \mathcal{G}_{(k+1)}\rangle + \frac{\rho_{(k)}}{2}\left\|\mathcal{Z}_{(k+1)} - \mathcal{G}_{(k+1)}\right\|_F^2,
\end{aligned} \tag{46}$$

and each term of (46) is nonnegative, we can deduce each term of (46) is bounded. The boundedness of $\left\|\mathcal{G}_{(k+1)}\right\|_{GNTR}$ implies that all singular values of $\mathcal{G}_{(k+1)}$ are bounded. Furthermore, based on the following equation

$$\left\|\mathcal{G}_{(k+1)}\right\|_F^2 = \frac{1}{n_3}\left\|\mathcal{G}_{f,k+1}\right\|_F^2 = \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{min(n_1,n_2)}(\mathcal{S}_f^i(j,j))^2, \tag{47}$$

we can derive the sequence $\{\mathcal{G}_{(k+1)}\}_{k=1}^{\infty}$ is bounded. Then, according to the updating formulas Eq. (14), Eq. (29), Eq. (28), it can be deduced that these sequences $\{\mathbf{Z}_{(k+1)}^v\}_{k=1}^{\infty}$, $\{\mathbf{A}_{(k+1)}\}_{k=1}^{\infty}$ and $\{\mathbf{P}_{(k+1)}^v\}_{k=1}^{\infty}$ are bounded.

Therefore, we can conclude that the sequence $\{\mathcal{L}_{1(k)} = (\mathbf{Z}_{(k)}^v, \mathbf{P}_{(k)}^v, \mathbf{A}_{(k)}, \mathbf{E}_{(k)}^v, \mathbf{Y}_{(k)}^v, \mathcal{W}_{(k)}, \mathcal{G}_{(k)})\}_{k=1}^{\infty}$ generated by Algorithm 1 is bounded.

2). Proof of $2nd$ part: According to Weierstrass-Bolzano theorem [71], there is at least one accumulation point of the sequence $\{\mathcal{L}_{1(k)}\}_{k=1}^{\infty}$, we denote one of the points as $\mathcal{L}_{1*}$. Then we obtain

$$\begin{aligned}&\lim_{k\to\infty}(\mathbf{Z}_{(k)}^v, \mathbf{P}_{(k)}^v, \mathbf{A}_{(k)}, \mathbf{E}_{(k)}^v, \mathbf{Y}_{(k)}^v, \mathcal{W}_{(k)}, \mathcal{G}_{(k)})\\ &= (\mathbf{Z}_*^v, \mathbf{P}_*^v, \mathbf{A}_*, \mathbf{E}_*^v, \mathbf{Y}_*^v, \mathcal{W}_*, \mathcal{G}_*).\end{aligned} \tag{48}$$

Form the updating rule of $\mathcal{W}$ and $\mathbf{Y}^v$, we have the following equations:

$$\begin{aligned}&\mathbf{X}^v - \mathbf{Z}_{anc(k+1)}^v\mathbf{A}_{(k+1)}\mathbf{P}_{(k+1)}^v - \mathbf{E}_{(k+1)}^v = (\mathbf{Y}_{(k+1)}^v - \mathbf{Y}_{(k)}^v)/\mu_{(k)},\\ &\mathcal{Z}_{(k+1)} - \mathcal{G}_{(k+1)} = (\mathcal{W}_{(k+1)} - \mathcal{W}_{(k)})/\rho_{(k)}.\end{aligned} \tag{49}$$

According to the boundedness of sequences $\{\mathcal{W}_{(k)}\}_{k=1}^{\infty}$ and $\{\mathbf{Y}_{(k)}^v\}_{k=1}^{\infty}$, we can obtain

$$\mathbf{X}^v - \mathbf{Z}_*^v\mathbf{A}_*\mathbf{P}_*^v - \mathbf{E}_*^v = 0, \quad \mathcal{Z}_* - \mathcal{G}_* = 0. \tag{50}$$

Furthermore, due to the first-order optimality conditions of $\mathbf{E}_{(k+1)}^v$ and $\mathcal{G}_{(k+1)}$, we can deduce:

$$\begin{aligned}&\mathbf{Y}_*^v \in \lambda\partial\left\|\mathbf{E}_*^v\right\|_{2,1}\\ &\mathcal{W}_* \in \partial\left\|\mathcal{G}_*\right\|_{GNTR}\end{aligned} \tag{51}$$

Thus, the accumulation point $\mathcal{L}_{1*}$ of sequence $\{\mathcal{L}_{1(k)}\}_{k=1}^{\infty}$ generated by Algorithm 1 satisfied the KKT condition. □

ACKNOWLEDGMENT

The authors would like to thank the reviewers and editor for their helpful comments to improve the paper. This work was supported by the Beijing Natural Science Foundation (No. 4242046).

REFERENCES

[1] J. Macqueen, "Some methods for classification and analysis of multivariate observations," *Proc. Symp. Math. Statist. and Probability, 5th*, vol. 1, 1967.
[2] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.
[3] G. Chao, S. Sun, and J. Bi, "A survey on multiview clustering," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 146–168, 2021.
[4] Y. Yang and H. Wang, "Multi-view clustering: A survey," *Big Data Mining and Analytics*, vol. 1, no. 2, pp. 83–107, 2018.
[5] L. Fu, P. Lin, A. V. Vasilakos, and S. Wang, "An overview of recent multi-view clustering," *Neurocomputing*, vol. 402, pp. 148–161, 2020.
[6] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, "Generalized latent multi-view subspace clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 1, pp. 86–99, 2018.
[7] H. Wang, Y. Yang, and B. Liu, "Gmc: Graph-based multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1116–1129, 2019.
[8] X. Li, H. Zhang, R. Wang, and F. Nie, "Multiview clustering: A scalable and parameter-free bipartite graph fusion method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 330–344, 2020.
[9] B. Zhang, Q. Qiang, F. Wang, and F. Nie, "Flexible multi-view unsupervised graph embedding," *IEEE Transactions on Image Processing*, vol. 30, pp. 4143–4156, 2021.
[10] H. Yang, Q. Gao, W. Xia, M. Yang, and X. Gao, "Multiview spectral clustering with bipartite graph," *IEEE Transactions on Image Processing*, vol. 31, pp. 3591–3605, 2022.
[11] W. Xia, X. Zhang, Q. Gao, X. Shu, J. Han, and X. Gao, "Multiview subspace clustering by an enhanced tensor nuclear norm," *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 8962–8975, 2022.
[12] Y. Jia, H. Liu, J. Hou, S. Kwong, and Q. Zhang, "Multi-view spectral clustering tailored tensor low-rank representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 12, pp. 4784–4797, 2021.
[13] Z. Lin, Z. Kang, L. Zhang, and L. Tian, "Multi-view attributed graph clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 2, pp. 1872–1880, 2021.
[14] M. White, X. Zhang, D. Schuurmans, and Y.-l. Yu, "Convex multi-view subspace learning," *Advances in Neural Information Processing Systems*, vol. 25, 2012.
[15] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 923–931.
[16] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 586–594.
[17] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4333–4341.
[18] S. Huang, Y. Liu, Y. Ren, I. W. Tsang, Z. Xu, and J. Lv, "Learning smooth representation for multi-view subspace clustering," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 3421–3429.
[19] P. Zhang, X. Liu, J. Xiong, S. Zhou, W. Zhao, E. Zhu, and Z. Cai, "Consensus one-step multi-view subspace clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 10, pp. 4676–4689, 2020.
[20] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, and Z. Xu, "Large-scale multi-view subspace clustering in linear time," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 4412–4419.
[21] M. Sun, P. Zhang, S. Wang, S. Zhou, W. Tu, X. Liu, E. Zhu, and C. Wang, "Scalable multi-view subspace clustering with unified anchors," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 3528–3536.
[22] S. Liu, S. Wang, P. Zhang, K. Xu, X. Liu, C. Zhang, and F. Gao, "Efficient one-pass multi-view subspace clustering with consensus anchors," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 7, 2022, pp. 7576–7584.
[23] B. Zhang, Q. Qiang, F. Wang, and F. Nie, "Fast multi-view semi-supervised learning with learned graph," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 286–299, 2020.
[24] X. Liu, L. Wang, X. Zhu, M. Li, E. Zhu, T. Liu, L. Liu, Y. Dou, and J. Yin, "Absent multiple kernel learning algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 6, pp. 1303–1316, 2019.
[25] J. Liu, X. Liu, Y. Yang, Q. Liao, and Y. Xia, "Contrastive multi-view kernel learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9552–9566, 2023.
[26] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *International Journal of Computer Vision*, vol. 126, no. 11, pp. 1157–1179, 2018.
[27] M.-S. Chen, C.-D. Wang, and J.-H. Lai, "Low-rank tensor based proximity learning for multi-view clustering," *IEEE Transactions*

on Knowledge and Data Engineering, vol. 35, no. 5, pp. 5076–5092, 2023.

[28] H. Lu, Q. Gao, Q. Wang, M. Yang, and W. Xia, "Centerless multi-view k-means based on the adjacency matrix," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 7, 2023, pp. 8949–8956.

[29] Z. Long, Q. Wang, Y. Ren, Y. Liu, and C. Zhu, "S2mvtc: a simple yet efficient scalable multi-view tensor clustering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 26 213–26 222.

[30] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1582–1590.

[31] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "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)*, 2016, pp. 5249–5257.

[32] P. Zhou, C. Lu, J. Feng, Z. Lin, and S. Yan, "Tensor low-rank representation for data recovery and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1718–1732, 2019.

[33] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 925–938, 2019.

[34] L. Fu, Z. Chen, Y. Chen, and S. Wang, "Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering," *IEEE Transactions on Multimedia*, vol. 25, pp. 4972–4985, 2023.

[35] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, "Logarithmic schatten-p norm minimization for tensorial multi-view subspace clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3396–3410, 2023.

[36] X. Sun, R. Zhu, M. Yang, X. Zhang, and Y. Tang, "Sliced sparse gradient induced multi-view subspace clustering via tensorial arctangent rank minimization," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[37] Y. Chen, S. Wang, C. Peng, Z. Hua, and Y. Zhou, "Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering," *IEEE Transactions on Image Processing*, vol. 30, pp. 4022–4035, 2021.

[38] W. Xia, Q. Gao, Q. Wang, X. Gao, C. Ding, and D. Tao, "Tensorized bipartite graph learning for multi-view clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 5187–5202, 2023.

[39] J. Ji and S. Feng, "Anchor structure regularization induced multi-view subspace clustering via enhanced tensor rank minimization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 343–19 352.

[40] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra and its Applications*, vol. 435, no. 3, pp. 641–658, 2011.

[41] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 1, pp. 148–172, 2013.

[42] Z. Ding and Y. Fu, "Robust multi-view subspace learning through dual low-rank decompositions," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, no. 1, 2016.

[43] M.-S. Chen, L. Huang, C.-D. Wang, D. Huang, and S. Y. Philip, "Multiview subspace clustering with grouping effect," *IEEE Transactions on Cybernetics*, vol. 52, no. 8, pp. 7655–7668, 2020.

[44] Y. Qin, N. Pu, and H. Wu, "Elastic multi-view subspace clustering with pairwise and high-order correlations," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 2, pp. 556–568, 2024.

[45] S. Wang, Y. Chen, L. Zhang, Y. Cen, and V. Voronin, "Hyper-laplacian regularized nonconvex low-rank representation for multi-view subspace clustering," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 8, pp. 376–388, 2022.

[46] Z. Kang, C. Peng, and Q. Cheng, "Robust pca via nonconvex rank approximation," in *2015 IEEE International Conference on Data Mining*, 2015, pp. 211–220.

[47] J. Lee and Y. Choe, "Low rank matrix recovery via augmented lagrange multiplier with nonconvex minimization," in *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 2016, pp. 1–5.

[48] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[49] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Transactions on Medical imaging*, vol. 28, no. 1, pp. 106–121, 2008.

[50] Y. Chen, S. Wang, X. Xiao, Y. Liu, Z. Hua, and Y. Zhou, "Self-paced enhanced low-rank tensor kernelized multi-view subspace clustering," *IEEE Transactions on Multimedia*, vol. 24, pp. 4054–4066, 2021.

[51] C. Peng, Z. Kang, M. Yang, and Q. Cheng, "Feature selection embedded subspace clustering," *IEEE Signal Processing Letters*, vol. 23, no. 7, pp. 1018–1022, 2016.

[52] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in *Proceedings of the 2003 American Control Conference, 2003.*, vol. 3. IEEE, 2003, pp. 2156–2162.

[53] S. Wang, X. Liu, S. Liu, J. Jin, W. Tu, X. Zhu, and E. Zhu, "Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences," *Advances in Neural Information Processing Systems*, vol. 35, pp. 5882–5895, 2022.

[54] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," *Advances in Neural Information processing Systems*, vol. 24, 2011.

[55] Y. Liu, X. Zhang, G. Tang, and D. Wang, "Multi-view subspace clustering based on tensor schatten-p norm," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 5048–5055.

[56] C. Lemaréchal and C. Sagastizábal, "Practical aspects of the moreau–yosida regularization: Theoretical preliminaries," *SIAM Journal on Optimization*, vol. 7, no. 2, pp. 367–385, 1997.

[57] P. D. Tao and L. H. An, "Convex analysis approach to dc programming: theory, algorithms and applications," *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.

[58] J. Ji and S. Feng, "High-order complementarity induced fast multi-view clustering with enhanced tensor rank minimization," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 328–336.

[59] Y. Liang, D. Huang, C.-D. Wang, and S. Y. Philip, "Multi-view graph learning by joint modeling of consistency and inconsistency," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022.

[60] G.-F. Lu, Q.-R. Yu, Y. Wang, and G. Tang, "Hyper-laplacian regularized multi-view subspace clustering with low-rank tensor constraint," *Neural Networks*, vol. 125, pp. 214–223, 2020.

[61] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[62] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," in *2004 Conference on Computer Vision and Pattern Recognition Workshop*, 2004, pp. 178–178.

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

[64] S. Liu, X. Liu, S. Wang, X. Niu, and E. Zhu, "Fast incomplete multi-view clustering with view-independent anchors," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

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

[66] M.-S. Chen, C.-D. Wang, D. Huang, J.-H. Lai, and P. S. Yu, "Efficient orthogonal multi-view subspace clustering," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 127–135.

[67] S. Liu, Q. Liao, S. Wang, X. Liu, and E. Zhu, "Robust and consistent anchor graph learning for multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–13, 2024.

[68] Y. Tang, Y. Xie, and W. Zhang, "Affine subspace robust low-rank self-representation: from matrix to tensor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9357–9373, 2023.

[69] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[70] A. S. Lewis and H. S. Sendov, "Nonsmooth analysis of singular values. part i: Theory," *Set-Valued Analysis*, vol. 13, no. 3, pp. 213–241, 2005.
[71] R. G. Bartle and D. R. Sherbert, *Introduction to real analysis*. Wiley New York, 2000, vol. 2.

Jintian Ji received the B.S. degree from Beijing Jiaotong University, China, in 2022. He is currently working toward the M.S. degree in the School of Computer Science and Technology, Beijing Jiaotong University. His main research interests include machine learning and computer vision, especially in multi-view clustering. He has published peer-reviewed papers in top-tier conferences including *ICCV, ACM SIGKDD, and ACM Multimedia*, respectively.

Songhe Feng received the B.S. and Ph.D degrees from the School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China, in 2003 and 2009, respectively. He is currently a full professor with the School of Computer Science and Technology, Beijing Jiaotong University. His research interests are machine learning and its applications, such as multi-view clustering, weakly-supervised multi-label learning, zero-shot learning and domain adaptation. He has published more than 60 peer-reviewed papers, including those in highly regarded journals and conferences such as the *IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Image Processing, IEEE Transactions on Cybernetics, IEEE Transactions on Multimedia, IEEE Transactions on Circuits and Systems for Video Technology, ACM Transactions on Knowledge Discovery from Data, ICML, NeurIPS, ICCV, AAAI, IJCAI, ACM SIGKDD, ACM Multimedia*, etc. He has been a visiting scholar with Concordia University, Canada and Michigan State University, USA in 2024 and 2014, respectively.