SLR-MVTC: Smooth Low-Rank Multi-View Tensor Clustering

Zhen Long, Yipeng Liu, Yazhou Ren, Ce Zhu

University of Electronic Science and Technology of China in Chengdu, 611731, China
{zhen.long,yipengliu,yazhou.ren,eczhu}@uestc.edu.cn

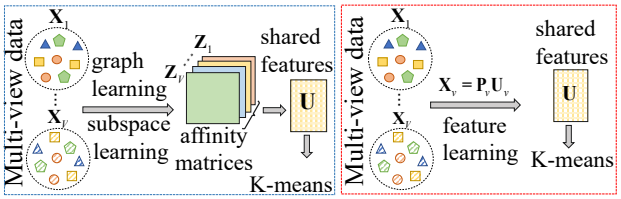
Abstract

Multi-view tensor clustering (MVTC) has gained much attention for its effectiveness in capturing global high-order correlations across views. However, current MVTC methods suffer from two limitations: 1) adopting a two-stage process to learn the latent features for clustering, and 2) either ignoring local similarities within views or treating local similarities and global high-order correlations equally. In this paper, we propose a smooth low-rank MVTC (SLR-MVTC) method, which aims to extract latent features that are smooth within each view and low-rank across views, enhancing clustering performance. Specifically, we first learn latent features from each view using orthogonal projection and then construct the latent feature tensor by concatenation and rotation. Then, we introduce a new smooth tensor nuclear norm to depict the low-rank components of the low-frequency parts in the feature tensor. Benefiting from the fast Fourier transform along the sample dimension, the obtained low-frequency components effectively capture local smoothness within views, while their low-rank parts further explore global correlations across views. Experimental results on six multi-view datasets demonstrate that SLR-MVTC outperforms state-of-the-art algorithms in terms of clustering performance and CPU time.

Code — https://github.com/longzhen520/SLR_MVTC

Introduction

Multi-view data collected from multiple feature extractors or sensors are ubiquitous in numerous real-world scenarios, with each view providing a distinct feature description (Zhang et al. 2019; Cui et al. 2023; Huang et al. 2021, 2023; Xu et al. 2024). For example, in assessing breast cancer risk, multi-view data include various types of ultrasound (US) images, such as US B-mode, US color Doppler, and US elastography images (Qian et al. 2021). Multi-view data can provide both consistency and complementary information, improving the performance of related data analysis tasks and leading to the development of multi-view learning (Zhang et al. 2024; Xu et al. 2023a; Yu et al. 2023). Among them, multi-view clustering (MVC) aims to divide data into several clusters by fully leveraging information from different views. It has garnered much attention in medical image segmentation (Zhou, Ruan, and Canu 2019), brain network analysis (Liu et al. 2018), and single-cell multi-omics integration (Huizing et al. 2023).



(a) Tensor based methods (b) MF based methods

Figure 1: Two ways for exploring inter-view correlations.

Existing MVC methods are mainly categorized into similarity learning-based and feature learning-based approaches based on the spaces in which they explore inter-view correlations. Similarity learning-based methods focus on capturing global high-order inter-view correlations by constructing affinity matrices. These matrices are typically derived through subspace learning or graph learning, which map the original data into sample spaces where the elements of the affinity matrix represent the similarity between pairs of samples. The final shared affinity matrix is then transformed into embedding features by performing eigenvalue decomposition on its Laplacian matrix, which is then used for K-means clustering (Khan and Maji 2019; Xia et al. 2022; Long et al. 2023), as shown in Fig. 1 (a). It can be observed this group employs a two-stage process to learn the latent features. Besides, one goal of clustering is to minimize the distance between samples within the same cluster, which means that samples should exhibit local similarities. Therefore, both high-order correlations across views and local similarities within each view are crucial. However, this group either ignores local similarities or treats local and global high-order correlations equally.

Another group focuses on directly learning consistent latent features from multi-view data for clustering, typically using deep leaning (Xu et al. 2023b; Ren et al. 2024; Yan et al. 2023) and matrix factorization (MF) (Liu et al. 2013; Wan et al. 2023; He et al. 2023; Liu et al. 2021). However, deep MVC requires substantial data and a meticulously designed network architecture. MF-based MVC maps the orig-

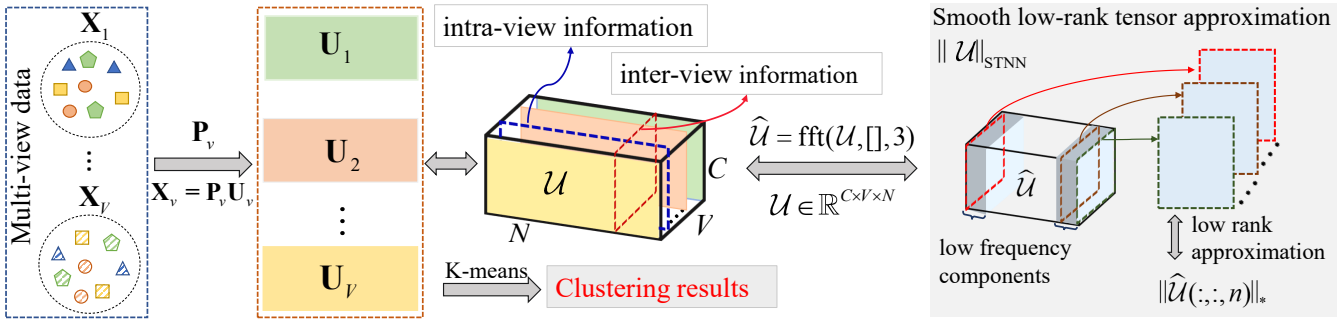


Figure 2: The framework of SLR-MVTC. This process involves first learning latent features $\mathbf{U}_v$ from the given data $\mathbf{X}_v$ using orthogonal projection $\mathbf{P}_v$. These latent features $\{\mathbf{U}_v\}_{v=1}^V$ are then formed into a tensor $\mathcal{U}$, where a newly defined smooth low-rank tensor approximation operator is applied to explore inter-view and intra-view correlations respectively.

inal features from different views into the consensus latent feature space for clustering, as shown in Fig. 1 (b). For instance, (Wang et al. 2017) proposed a diverse nonnegative MF to enhance the diversity between latent features. (Wan et al. 2023) fused the latent features by mapping them into a consensus low-dimensional space using rotation matrices. However, current MF-based MVC methods explore pairwise correlations between views, ignoring higher-order correlations across views and local smoothness within views.

To address these issues, we propose a smooth low-rank multi-view tensor clustering (SLR-MVTC) method, designed to efficiently explore high-order inter-view and local smooth intra-view correlations of latent features respectively, as shown in Fig. 2. In particular, given multi-view data $\{\mathbf{X}_v \in \mathbb{R}^{D_v\times N}\}_{v=1}^V$, we first obtain latent features $\{\mathbf{U}_v \in \mathbb{R}^{C\times N}\}_{v=1}^V$ using orthogonal projection matrices $\{\mathbf{P}_v \in \mathbb{R}^{D_v\times C}\}_{v=1}^V$. Next, to respectively capture intra-view local similarities and inter-view global correlations of latent features, we introduce a new norm $\|\mathcal{U}\|_{\text{STNN}}$ on the feature tensor $\mathcal{U} \in \mathbb{R}^{C\times V\times N}$. Here $\mathcal{U}$ is formed by concatenating $\mathbf{U}_v$ along the third dimension and then rotating it.

The norm $\|\mathcal{U}\|_{\text{STNN}}$ is defined based on the work (Long et al. 2024), which introduced tensor low-frequency operators to obtain smooth representations of samples. The newly defined norm advances the approach by further depicting the low-rank component of the low-frequency parts. In this way, the low-frequency component provides a smooth representation within views, and its low-rank parts further explore the global correlations across views. Finally, the shared features $\mathbf{U} = \frac{1}{V}\sum_{v=1}^V \mathbf{U}_v$ are fed into the K-means algorithm for clustering. Experimental results on six multi-view datasets show that SLR-MVTC outperforms state-of-the-art algorithms in both clustering performance and computational efficiency.

Our contributions beyond existing MVC methods are:

- We develop the norm $\|\mathcal{U}\|_{\text{STNN}}$, which captures both global correlations across views and local smoothness within views, with a computational complexity of $O(N \log N)$.
- We integrate the newly defined norm with latent feature learning into a unified framework, allowing the smooth low-rank prior to guiding the learning of latent features from multi-view data during each iteration, resulting in enhanced clustering performance.
- Experimental results on six multi-view datasets show that SLR-MVTC outperforms state-of-the-art algorithms in both clustering performance and CPU time.

Notations and Problem Formulation

Notations

For clarity, we present frequently used notations in Table 1.

Symbols	Descriptions
a, $\mathbf{a}$, $\mathbf{A}$, $\mathcal{A}$	scalar, vector, matrix, tensor
$\|\mathbf{A}\|_{\text{F}}$	$\sqrt{\sum_{i=1}^I \sum_{j=1}^J a_{i,j}^2}$
$\|\mathbf{A}\|_*$	$\sum_{i=1}^I \sigma_i$, σ_i is the i-th singular value of $\mathbf{A}$
$\|\mathbf{A}\|_{2,1}$	$\sum_{i=1}^I \|\mathbf{a}\|_2$
trace($\mathbf{A}$)	$\sum_{i=1}^I a_{i,i}$
$\lfloor a \rfloor$	floor function
N, V, C	number of samples, views, clusters
$v = 1, \cdots, V$	index ranges from 1 to its capital version
$\mathbf{X}_v \in \mathbb{R}^{D_v\times N}$	multi-view data in the v-th view
$\mathbf{P}_v \in \mathbb{R}^{D_v\times C}$	orthogonal projection matrix in the v-th view
$\mathbf{U}_v \in \mathbb{R}^{C\times N}$	latent features in the v-th view
$\mathcal{U} \in \mathbb{R}^{C\times V\times N}$	latent feature tensor

Table 1: Frequently used notations in this paper.

Preliminaries

Definition 1. *(t-SVD) (Kilmer et al. 2013; Lu et al. 2020; Braman 2010) Given a tensor $\mathcal{U} \in \mathbb{R}^{C\times V\times N}$, its t-SVD can be expressed as:*

$$\mathcal{U} = \mathcal{S} * \mathcal{V} * \mathcal{D}^{\text{T}}, \tag{1}$$

where $\mathcal{S} \in \mathbb{R}^{C\times C\times N}$ and $\mathcal{D} \in \mathbb{R}^{V\times V\times N}$ are orthogonal tensors, and $\mathcal{V} \in \mathbb{R}^{C\times V\times N}$ is an f-diagonal tensor.

The t-SVD can be obtained by Algorithm 1, where $\widehat{\mathcal{U}} = \text{fft}(\mathcal{U}, [\ ], 3)$ denotes applying the fast Fourier transform (FFT) operator along the 3-rd dimension, and ifft denotes the inverse FFT.

Algorithm 1: t-SVD

Input: $\mathcal{U} \in \mathbb{R}^{C\times V\times N}$
Output: $\mathcal{S} \in \mathbb{R}^{C\times C\times N}$, $\mathcal{V} \in \mathbb{R}^{C\times V\times N}$, $\mathcal{D} \in \mathbb{R}^{V\times V\times N}$;
$\widehat{\mathcal{U}} \leftarrow \text{fft}(\mathcal{U}, [\,], 3)$;
for $n = 1$ to N **do**
  $[\mathbf{S}, \mathbf{V}, \mathbf{D}] = \textbf{svd}(\widehat{\mathcal{U}}(:, :, n))$;
  $\widehat{\mathcal{S}}(:, :, n) = \mathbf{S}; \widehat{\mathcal{V}}(:, :, n) = \mathbf{V}; \widehat{\mathcal{D}}(:, :, n) = \mathbf{D}$.
end for
$\mathcal{S} = \text{ifft}(\widehat{\mathcal{S}}, [\,], 3); \mathcal{V} = \text{ifft}(\widehat{\mathcal{V}}, [\,], 3); \mathcal{D} = \text{ifft}(\widehat{\mathcal{D}}, [\,], 3)$.

Definition 2. *(**Tensor nuclear norm (TNN)**) (Kilmer and Martin 2011; Xie et al. 2021) Given a tensor $\mathcal{U} \in \mathbb{R}^{C\times V\times N}$, its t-SVD-based TNN is defined as*

$$\|\mathcal{U}\|_* = \sum_{n=1}^{N} \|\widehat{\mathcal{U}}(:, :, n)\|_* = \sum_{n=1}^{N} \sum_{j=1}^{\min(C,V)} \sigma_j(\widehat{\mathcal{U}}(:, :, n)), \quad (2)$$

where $\sigma_j(\cdot)$ is the j-th largest singular value of $\widehat{\mathcal{U}}(:, :, n)$, $(n = 1, 2, \ldots, N)$ and $\widehat{\mathcal{U}}(:, :, n)$ is the n-th frontal slice of $\widehat{\mathcal{U}}$.

Definition 3. *(**Tensor low-frequency component (TLFC)**) (Long et al. 2024) The TLFC of a tensor $\mathcal{U} \in \mathbb{R}^{C\times V\times N}$ is defined as the Frobenius norm of the frontal slices in the low-frequency domain. Mathematically, it can be expressed as follows:*

$$\|\mathcal{U}\|_{\text{TLFC}} = \|\widehat{\mathcal{U}}(:, :, 1) + \sum_{n=2}^{L} (\widehat{\mathcal{U}}(:, :, n) + \widehat{\mathcal{U}}(:, :, N + 2 - n))\|_{\text{F}}, \quad (3)$$

where L is the number of low-frequency bands, $\widehat{\mathcal{U}}(:, :, 1)$ is the 0-frequency component and $\widehat{\mathcal{U}}(:, :, N+2-n) = \text{conj}(\widehat{\mathcal{U}}(:, :, n))$ means conjugate symmetry according to the properties of the FFT.

Related Works

MF-based MVC MF-based MVC methods seek to discover the common features $\mathbf{U}$ that reveal the consensus structure from the given multi-view data $\{\mathbf{X}_v \in \mathbb{R}^{D^v\times N}\}_{v=1}^{V}$. The general framework is described as follows:

$$\min_{\mathbf{P}_v\geq 0, \mathbf{U}_v\geq 0} \sum_{v=1}^{V} \|\mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v\|_{\text{F}} + \lambda \sum_{v=1}^{V} \Re(\mathbf{U}, \mathbf{U}_v), \quad (4)$$

where $\Re(\mathbf{U}, \mathbf{U}_v)$ is the regularization terms. For instance, (Liu et al. 2013) used $\|\mathbf{U} - \mathbf{U}_v\|_{\text{F}}$ to determine the common features. Furthermore, (Wang et al. 2017) introduced a diversity constraint, $\text{trace}(\mathbf{U}_v\mathbf{U}_w^{\text{T}})$, to promote orthogonality between $\mathbf{U}_v$ and $\mathbf{U}_w$. Besides, some approaches have removed the non-negativity constraint on $\mathbf{P}_v$ and $\mathbf{U}_v$, and instead introduce orthogonality regularization for $\mathbf{P}_v$ to get more distinctive embedding features. The approach is as follows:

$$\min_{\mathbf{P}_v, \mathbf{U}_v} \sum_{v=1}^{V} \|\mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v\|_{\text{F}}, \text{ s. t. } \mathbf{P}_v\mathbf{P}_v^{\text{T}} = \mathbf{I}_C. \quad (5)$$

According to this framework, (Liu et al. 2021) further decomposes the features $\mathbf{U}_v$ into a consensus indicator matrix and centroid matrix. Besides, (Wan et al. 2023) maps the original feature matrix of each view into several latent features and dynamically adjusts their weights based on their corresponding contributions.

The correlations in multi-view data ($V > 2$) are of a higher order. However, this group only considers pairwise correlations between views and do not explore the higher-order correlations across views .

Tensor-based MVC Tensor-based MVC methods aim to identify the common affinity matrix by capturing high-order correlations of memberships between different views. Given multi-view data $\{\mathbf{X}_v \in \mathbb{R}^{D^v\times N}\}_{v=1}^{V}$, the general model can be formulated as follows:

$$\min_{\mathbf{E}_v, \mathbf{Z}_v} \mathfrak{L}(\mathcal{Z}) + \lambda\Re(\mathbf{E}), \text{ s. t. } \mathbf{X}_v = f(\mathbf{Z}_v) + \mathbf{E}_v, \quad (6)$$

where $\mathbf{Z}_v$ refers to the relationship matrix, and $\mathcal{Z} = \Omega(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_V)$ means the combination of these relationship matrices into a third-order tensor. $\mathbf{E} = [\mathbf{E}_1; \mathbf{E}_2; \ldots; \mathbf{E}_V]$ is the noise. $\mathfrak{L}(\mathcal{Z})$ and $\Re(\mathbf{E})$ are the regularization terms.

When $f(\mathbf{Z}_v) = \mathbf{X}_v\mathbf{Z}_v$, the problem (6) can be transformed into the optimization problem for subspace learning. For instance, (Xie et al. 2018) first considered t-SVD-based TNN on the rotated self-representation tensor to explore high-order correlations across views. Furthermore, some low-rank tensor network approximations are considered on the self-representation tensor to characterize the intra-view and inter-view relationships (Lu et al. 2023; Liu et al. 2024).

When $f(\mathbf{Z}_v) = \mathbf{A}_v\mathbf{Z}_v$, the problem (6) is the optimization problem of anchor learning. For example, (Long et al. 2023) considered low-rank MERA approximation on the anchor graph tensor to capture the inter/intra-view correlations. (Ji and Feng 2023) split the anchor graph tensor into two parts: the common and the specific. To investigate the common and specific, respectively, enhanced tensor rank and tensorial exclusive regularization are taken into consideration. Furthermore, (Long et al. 2024) investigated the learning of embedding features from the provided anchor graph and introduced a low-frequency operator to obtain a smooth representation of the samples. (Li et al. 2023) introduced a method to directly learn the common indicator feature using orthogonal non-negative tensor factorization on the anchor graph tensor.

This group employs a two-stage processing approach to acquire the final embedding features. Besides, the aforementioned methods either ignore local similarities or treat local similarity and global high-order correlations equally, failing to effectively explore multi-view data for clustering.

Proposed Method

In this paper, we focus on directly learning the latent features from multi-view data and introduce a smooth low-rank tensor approximation method on the feature tensor to capture the local similarity within views and the global high-order correlations across views.

Model Development

Given $\{\mathbf{X}_v \in \mathbb{R}^{D_v \times N}\}_{v=1}^V$, the general optimization problem of SLR-MVTC is expressed as:

$$\begin{aligned} &\min_{\{\mathbf{E}_v,\mathbf{U}_v,\mathbf{P}_v\}_{v=1}^V} \gamma\|\mathcal{U}\|_{\text{STNN}} + \lambda\sum_{v=1}^{V}\|\mathbf{E}_v\|_{2,1} \\ &\text{s. t. } \mathbf{X}_v = \mathbf{P}_v\mathbf{U}_v + \mathbf{E}_v, \mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v = \mathbf{I}_C, v = 1, \cdots, V. \end{aligned} \tag{7}$$

Here, the orthogonal projection matrix $\mathbf{P}_v \in \mathbb{R}^{D_v \times C}$ is used to obtain more distinctive latent features $\mathbf{U}_v \in \mathbb{R}^{C\times N}$, where C is the number of clusters. In addition, $\mathbf{E}_v$ denotes the sparse noise, which is typically removed using convex surrogates such as ℓ_1 or $\ell_{2,1}$ norms. $\mathcal{U} = \Omega(\mathbf{U}_1, \cdots, \mathbf{U}_V) \in \mathbb{R}^{C\times V\times N}$ is the latent feature tensor, where Ω is an operator that stacks all latent features into a 3-rd tensor and subsequently rotates it. The inverse operator is denoted as $\mathbf{U}_v = \Omega_v^{-1}(\mathcal{U})$.

$\|\mathcal{U}\|_{\text{STNN}}$ is the newly defined norm, which aims to explore the global high-order inter-view correlations and local intra-view similarities. According to definitions of TNN and TLFC, we introduce the definition of STNN as follows:

Definition 4. *(Smooth tensor nuclear norm (STNN)) Given a tensor $\mathcal{U} \in \mathbb{R}^{C\times V\times N}$, its STNN is defined as the nuclear norm of the frontal slices in the low-frequency part. Mathematically, it can be expressed as follows:*

$$\|\mathcal{U}\|_{\text{STNN}} = \|\widehat{\mathcal{U}}(:,:,1)\|_* + \sum_{n=2}^{L}\|\widehat{\mathcal{U}}(:,:,n)\|_* + \|\widehat{\mathcal{U}}(:,:,N+2-n)\|_*, \tag{8}$$

where $\|\widehat{\mathcal{U}}(:,:,1)\|_*$ is the nuclear norm of 0-frequency component and L represents the number of low-frequency bands.

Note that the low-frequency components obtained using the FFT along the sample dimension will result in a local smooth representation. Furthermore, the low-rank approximation in the low-frequency part will capture the global high-order correlations.

To make the optimization problem (7) separable, we introduce an auxiliary variable $\mathcal{Y}$, leading to the following reformulation:

$$\begin{aligned} &\min_{\{\mathbf{E}_v,\mathbf{U}_v,\mathbf{P}_v\}_{v=1}^V,\mathcal{Y}} \gamma\|\mathcal{Y}\|_{\text{STNN}} + \lambda\sum_{v=1}^{V}\|\mathbf{E}_v\|_{2,1} \\ &\text{s. t. } \mathbf{X}_v = \mathbf{P}_v\mathbf{U}_v + \mathbf{E}_v, \mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v = \mathbf{I}_C, v = 1, \cdots, V, \mathcal{U} = \mathcal{Y}. \end{aligned} \tag{9}$$

Solutions

To solve the problem with the constraints mentioned above, the Alternating Direction Method of Multipliers (ADMM) (Boyd et al. 2011) framework can be used. The corresponding augmented Lagrangian function can be defined as follows:

$$\begin{aligned} &\mathfrak{L}\left(\{\{\mathbf{U}_v, \mathbf{E}_v, \mathbf{P}_v, \mathbf{Q}_v\}_{v=1}^V, \mathcal{Y}, \mathcal{F}\right) \\ &= \sum_{v=1}^{V}(\langle\mathbf{Q}_v, \mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v - \mathbf{E}_v\rangle + \frac{\mu_1}{2}\|\mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v - \mathbf{E}_v\|_{\mathrm{F}}^2 \\ &+ \lambda\|\mathbf{E}_v\|_{2,1}) + \gamma\|\mathcal{Y}\|_{\text{STNN}} + \langle\mathcal{F}, \mathcal{U} - \mathcal{Y}\rangle + \frac{\mu_2}{2}\|\mathcal{U} - \mathcal{Y}\|_{\mathrm{F}}^2, \end{aligned} \tag{10}$$

Algorithm 2: Updating $\mathcal{Y}$

Input: $\mathcal{H}$, L, γ
Initialize: $\mathcal{H} = (\mathcal{U} + \frac{\mathcal{F}}{\mu_2})$, $\tau = \frac{\gamma}{\mu_2}$.
Output: $\mathcal{Y}$;
$\widehat{\mathcal{H}} \leftarrow \text{fft}(\mathcal{H}, [\,], 3)$;
$\widehat{\mathcal{Y}}(:,:,1)) = \text{SVT}_\tau(\widehat{\mathcal{H}}(:,:,1))$;
for $n = 2$ to L **do**
  $\widehat{\mathcal{Y}}(:,:,n) = \text{SVT}_\tau(\widehat{\mathcal{H}}(:,:,n))$;
  $\widehat{\mathcal{Y}}(:,:,N+2-n) = \text{conj}(\widehat{\mathcal{Y}}(:,:,n))$.
end for
$\mathcal{Y} = \text{ifft}(\widehat{\mathcal{Y}}, [\,], 3)$.

under constraints $\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v = \mathbf{I}_C, v = 1, \cdots, V$, where $\{\mathbf{Q}_v\}_{v=1}^V$, $\mathcal{F}$ are Lagrange multipliers and μ_1, μ_2 are penalty factors. Under the ADMM framework, problem (10) can be divided into several subproblems, where each subproblem alternately updates one variable while keeping the others fixed.

Update $\{\mathbf{P}_v\}_{v=1}^V$: Fixing other variables, the subproblem of $\mathbf{P}_v$ can be rewritten as:

$$\max_{\mathbf{P}_v:\mathbf{P}_v^{\mathrm{T}}\mathbf{P}_v=\mathbf{I}_C} \text{trace}(\mathbf{P}_v(\mathbf{U}_v(\mathbf{Q}_v + \mu_1\mathbf{X}_v - \mu_1\mathbf{E}_v)^{\mathrm{T}})). \tag{11}$$

This subproblem is a well-known orthogonal Procrustes problem. Letting $\mathbf{H} = \mathbf{U}_v(\mathbf{Q}_v+\mu_1\mathbf{X}_v-\mu_1\mathbf{E}_v)^{\mathrm{T}}$, and $[\mathbf{S}, \mathbf{V}, \mathbf{D}] = \textbf{svd}(\mathbf{H})$ is the singular value decomposition of $\mathbf{H}$, the solution $\mathbf{P}_v = \mathbf{D}\mathbf{S}^{\mathrm{T}}$. The computational complexity of updating $\mathbf{P}_v$ is $O(C^2D_v)$, where D_v is the original feature dimension.

Update $\{\mathbf{U}_v\}_{v=1}^V$: The subproblem of $\mathbf{U}_v$ can be written as:

$$\begin{aligned} \min_{\mathbf{U}_v}\ &\langle\mathbf{Q}_v, \mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v - \mathbf{E}_v\rangle + \frac{\mu_1}{2}\|\mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v - \mathbf{E}_v\|_{\mathrm{F}}^2 \\ &+ \langle\mathcal{F}, \mathcal{U} - \mathcal{Y}\rangle + \frac{\mu_2}{2}\|\mathcal{U} - \mathcal{Y}\|_{\mathrm{F}}^2. \end{aligned} \tag{12}$$

By taking the derivative and setting it to zero, the solution of $\mathbf{U}_v$ can be obtained as follows:

$$\mathbf{U}_v = \frac{\mathbf{P}_v^{\mathrm{T}}(\mathbf{Q}_v + \mu_1\mathbf{X}_v - \mu_1\mathbf{E}_v) + \Omega_v^{-1}(\mu_2\mathcal{Y} - \mathcal{F})}{\mu_1 + \mu_2}, \tag{13}$$

where $\Omega_v^{-1}(\mathcal{Y}) = \mathbf{Y}_v$. The computational complexity of solving $\mathbf{U}_v$ is $O(NCD_v)$.

Update $\{\mathbf{E}_v\}_{v=1}^V$**:** The solution of $\mathbf{E}_v$ can be updated by

$$\mathbf{E}_v = \text{sth}(\mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v + (1/\mu_1)\mathbf{Q}_v, \lambda/\mu_1), \tag{14}$$

where $\text{sth}(x, \tau)$ is the well-known soft thresholding operator, defined as follows: $\text{sth}(x, \tau) = \text{sgn}(x)\max(|x| - \tau, 0)$. The computational complexity of solving $\mathbf{E}_v$ is $O(NCD_v)$.

Update $\mathcal{Y}$: The subproblem of $\mathcal{Y}$ is rewritten as:

$$\min_{\mathcal{Y}} \frac{\gamma}{\mu_2}\|\mathcal{Y}\|_{\text{STNN}} + \frac{1}{2}\|\mathcal{Y} - (\mathcal{U} + \frac{\mathcal{F}}{\mu_2})\|_{\mathrm{F}}^2, \tag{15}$$

which can be solved by Algorithm 2.

Here $\text{SVT}_\tau(\mathbf{H})$ is the well-known singular value thresholding operator, defined as follows:

$$\text{SVT}_\tau(\mathbf{H}) = \mathbf{S}\,\text{sth}(\mathbf{V}, \tau)\mathbf{D}^{\mathrm{T}},$$

Algorithm 3: SLR-MVTC

Input: Multi-view data $\{\mathbf{X}_v\}_{v=1}^V$, low frequency parameter L,low rank parameter γ, regularization parameter λ
Initialize: $\mu_1 = 5 \times 10^{-4}$, $\mu_2 = 10^{-3}$
while not converged **do**
  for $v = 1$ to V **do**
    Update $\mathbf{P}_v$ via Eq.11;
    Update $\mathbf{U}_v$ via Eq.13
    Update $\mathbf{E}_v$ via Eq.14;
    Update $\mathbf{Q}_v$ via Eq.16;
  end for
  Update $\mathcal{Y}$ via Alg.2;
  Update $\mathcal{F}$ via Eq. 17;
  $\mu_1 = \min(\mu_1 * 1.5, 10^{10})$, $\mu_2 = \min(\mu_2 * 1.5, 10^{10})$;
  $L = \lfloor \min(L * 1.5, N/2) \rfloor$;
end while
Apply K-means on $\mathbf{U} = \frac{1}{V}\sum_{v=1}^V \mathbf{U}_v$
Output: Clustering result.

where $[\mathbf{S}, \mathbf{V}, \mathbf{D}] = \mathbf{svd}(\mathbf{H})$. The computational complexity of updating $\mathcal{Y}$ is $O(\max(N \log N, V^2 C))$, where $O(N \log N)$ is from the FFT and $O(V^2 C)$ comes from the SVD operator.

Update Lagrangian multipliers:

$$\mathbf{Q}_v = \mathbf{Q}_v + \mu_1(\mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v - \mathbf{E}_v), v = 1, \cdots, V, \quad (16)$$

$$\mathcal{F} = \mathcal{F} + \mu_2(\mathcal{U} - \mathcal{Y}). \quad (17)$$

Overall, the solution of SLR-MVTC is summarized in Algorithm 3, with convergence achieved when RE $\leq 10^{-5}$, where RE=$\max(\max_v(\|\mathbf{X}_v - \mathbf{P}_v\mathbf{U}_v - \mathbf{E}_v\|_\infty), \|\mathcal{U} - \mathcal{Y}\|_\infty)$. Besides, the main computational complexity of SLR-MVTC is $O(\max(N \log N, NC \sum_{v=1}^V D_v))$.

Experiments

Experimental Settings

Multi-View Datasets Description Six well-known multi-view datasets were selected to evaluate the performance of SLR-MVTC: ORL (Samaria and Harter 1994), CCV (Jiang et al. 2011), ALOI_100 (Geusebroek, Burghouts, and Smeulders 2005), Reuters (Lewis et al. 2004), AwA (Lampert, Nickisch, and Harmeling 2009), and CIFAR100 (Krizhevsky, Hinton et al. 2009). Detailed statistical information about these datasets is provided in Table 2.

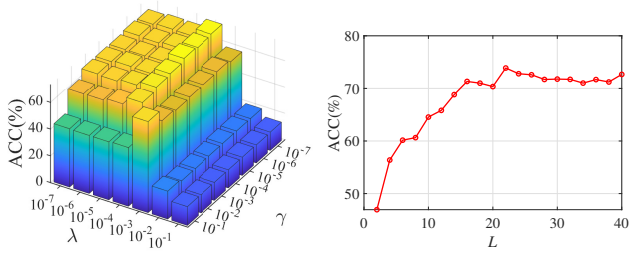
Datasets	N	V	C	$(D_1, \cdots, D_V)$
ORL	400	3	40	(4096, 3304, 6750)
CCV	6773	3	20	(20, 20, 20)
ALOI_100	10800	4	100	(77, 13, 64, 125)
Reuters	18758	5	6	(21531, 24892, 34251, 15506, 11547)
AwA	30475	6	50	(2688, 2000, 252, 2000, 2000)
CIFAR100	50000	3	100	(512,2048,1024)

Table 2: The statistical information of multi-view datasets, where N, V, C are the number of samples, views, and clusters, respectively. $(D_1, \cdots, D_V)$ are original feature sizes.

Compared Clustering Algorithms To evaluate the clustering performance, we compare the proposed approach against eight state-of-the-art clustering methods. These include four pairwise inter-view correlation-based methods: Binary Multi-View Clustering (BMVC) (Zhang et al. 2018), Fast Parameter-free Multi-view Subspace Clustering with Consensus Anchor Guidance (FPMSC-CAG) (Wang et al. 2022), and Auto-Weighted Multi-View Clustering for Large-Scale Data (AWMVC) (Wan et al. 2023), A Simple yet Efficient Scalable Multi-View Tensor Clustering (S^2MVTC) (Long et al. 2024). Additionally, we compare three high-order inter-view correlation-based multi-view clustering methods: scalable MERA based multi-view clustering (sMERA-MVC) (Long et al. 2023), Orthogonal Non-negative Tensor Factorization-based Multi-view Clustering (Orth-NTF) (Li et al. 2023), and High-order Complementarity Induced Fast Multi-View Clustering with Enhanced Tensor Rank Minimization (CFMVC-ETR) (Ji and Feng 2023). We also add one updated multi-view clustering method: Efficient Balanced Multi-view Graph Clustering via Good Neighbor Fusion (EBMGC-GNF) (Wu et al. 2024). All tests are tuned best and accomplished on a desktop computer with a 2.10 GHz 13th Gen Intel(R) Core(TM) i7 Processor, 64 GB RAM, and MATLAB 2020b.

Evaluation Metrics The performance of the clustering methods is evaluated using seven standard metrics: Accuracy (ACC), Normalized Mutual Information (NMI), F-score, Precision, Recall, Adjusted Rand Index (ARI), and CPU time. For all metrics except CPU time, higher values indicate better clustering performance.

Parameter Settings SLR-MVC has three parameters—λ, γ, and L—which control the weight on noise, the low-rank component, and the low-frequency component, respectively. We select these parameters by a brute-force search, with λ and γ ranging from $\{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ and L ranging from {2:2:40}. Fig. 3 shows partial results with some parameters fixed. For example, with $L = 22$, the ranges of λ and γ are shown in Fig. 3 (a), where ACC performs better and remains stable with smaller γ and $\lambda = 10^{-4}$. Additionally, Fig. 3 (b) demonstrates that when $L > 16$, the ACC shows little variation and remains above 70%.



(a) $L = 22$ (b) $\gamma = 10^{-5}, \lambda = 10^{-4}$

Figure 3: The change of clustering performance on CCV dataset. (a) γ and λ changed with $L = 22$; (b) L changed with $\gamma = 10^{-5}, \lambda = 10^{-4}$.

Datasets	Methods	F-score	Precision	Recall	NMI	ARI	ACC	CPU time (s)
ORL	BMVC (TPAMI'18)	59.61(0.00)	55.12(0.00)	64.89(0.00)	84.32(0.00)	58.60(0.00)	71.25(0.00)	0.08
	AWMVC (AAAI'23)	66.61(2.87)	62.39(3.32)	71.48(2.71)	87.80(1.16)	65.78(2.95)	73.88(2.24)	6.41
	FPMVS-CAG (TIP'22)	57.74(2.24)	50.91(2.95)	66.78(1.10)	83.84(0.70)	56.63(2.32)	67.63(2.12)	26.33
	S^2MVTC (CVPR'24)	76.40(0.00)	67.43(0.00)	88.11(0.00)	92.49(0.00)	75.78(0.00)	81.50(0.00)	0.08
	EBMGC-GNF (TPAMI'24)	80.64(0.00)	79.85(0.00)	81.44(0.00)	92.76(0.00)	80.19(0.00)	85.50(0.00)	1.26
	sMERA-MVC (TMM'24)	85.92(3.56)	81.52(4.58)	90.88(2.39)	95.67(1.09)	85.58(3.65)	88.38(3.06)	4.22
	CFMVC-ETR (ACMMM'23)	90.10(0.00)	84.80(0.00)	**96.11(0.00)**	**97.74(0.00)**	89.86(0.00)	90.00(0.00)	5.47
	Orth-NTF (ICCV'24)	59.04(0.00)	57.29(0.00)	60.89(0.00)	84.57(0.00)	58.06(0.00)	72.50(0.00)	2.09
	SLR-MVC	**91.37(1.66)**	**88.62(2.09)**	94.29(1.33)	97.30(0.56)	**91.16(1.70)**	**92.90(1.49)**	2.10
CCV	BMVC (TPAMI'18)	9.48(0.00)	9.55(0.00)	9.68(0.00)	9.80(0.00)	4.02(0.00)	15.44(0.00)	1.25
	AWMVC (AAAI'23)	11.29(0.09)	11.86(0.11)	10.77(0.08)	14.59(0.18)	6.17(0.10)	18.85(0.17)	0.40
	FPMVS-CAG (TIP'22)	12.99(0.18)	13.62(0.24)	12.42(0.15)	18.37(0.26)	7.96(0.21)	22.13(0.34)	7.36
	S^2MVTC (CVPR'24)	48.58(0.00)	40.54(0.00)	60.60(0.00)	60.87(0.00)	44.79(0.00)	58.39(0.00)	0.97
	EBMGC-GNF (TPAMI'24)	10.45(0.00)	11.23(0.00)	9.78(0.00)	14.06(0.00)	5.41(0.00)	17.27(0.00)	9.51
	sMERA-MVC (TMM'24)	59.85(3.21)	59.54(4.99)	60.29(1.85)	71.24(1.07)	57.37(3.49)	67.84(3.52)	2.09
	CFMVC-ETR (ACMMM'23)	60.19(0.00)	60.29(0.00)	60.10(0.00)	**75.49(0.00)**	57.76(0.00)	68.96(0.00)	3.42
	Orth-NTF (ICCV'24)	26.37(0.00)	27.74(0.00)	25.13(0.00)	46.25(0.00)	22.10(0.00)	35.43(0.00)	54.07
	SLR-MVC	**65.53(1.70)**	**65.28(1.92)**	**65.80(1.78)**	74.54(0.53)	**63.42(1.81)**	**72.84(2.07)**	0.72
ALOI_100	BMVC (TPAMI'18)	1.96(0.00)	0.99(0.00)	**100.00(0.00)**	0.00(0.00)	0.00(0.00)	1.00(0.00)	3.29
	AWMVC (AAAI'23)	50.24(1.25)	46.92(1.79)	54.08(0.64)	75.27(0.45)	49.71(1.27)	61.73(1.31)	1.63
	FPMVS-CAG (TIP'22)	60.88(1.45)	56.27(1.96)	66.32(0.95)	83.22(0.41)	60.45(1.47)	69.31(1.51)	16.25
	S^2MVTC (CVPR'24)	52.92(0.00)	38.09(0.00)	86.66(0.00)	86.59(0.00)	52.26(0.00)	50.81(0.00)	1.74
	EBMGC-GNF (TPAMI'24)	71.52(0.00)	71.16(0.00)	71.87(0.00)	86.38(0.00)	71.23(0.00)	81.95(0.00)	17.00
	sMERA-MVC (TMM'24)	71.96(1.34)	69.96(1.45)	74.08(1.25)	90.81(0.39)	71.67(1.35)	78.42(1.46)	5.73
	CFMVC-ETR (ACMMM'23)	68.18(0.00)	61.01(0.00)	77.25(0.00)	89.80(0.00)	67.82(0.00)	74.33(0.00)	31.02
	Orth-NTF (ICCV'24)	33.29(0.00)	30.13(0.00)	37.19(0.00)	73.68(0.00)	32.55(0.00)	45.29(0.00)	1229.19
	SLR-MVC	**81.89(1.19)**	**79.09(1.60)**	84.90(0.80)	**93.89(0.26)**	**81.70(1.20)**	**85.49(1.51)**	7.14
Reuters	BMVC (TPAMI'18)	47.28(0.00)	51.33(0.00)	43.83(0.00)	33.12(0.00)	34.34(0.00)	56.69(0.00)	6.03
	AWMVC (AAAI'23)	38.33(0.00)	33.90(0.00)	44.11(0.00)	25.59(0.00)	18.65(0.00)	45.86(0.00)	6082.89
	FPMVS-CAG (TIP'22)	-	-	-	-	-	-	-
	S^2MVTC (CVPR'24)	95.86(0.00)	95.91(0.00)	95.81(0.00)	92.23(0.00)	94.73(0.00)	96.50(0.00)	5.59
	EBMGC-GNF (TPAMI'24)	38.01(0.00)	41.64(0.00)	34.96(0.00)	30.22(0.00)	22.96(0.00)	47.60(0.00)	9973.23
	sMERA-MVC (TMM'24)	90.25(1.36)	91.61(0.41)	88.95(2.26)	89.47(0.87)	87.64(1.69)	88.13(1.42)	2513.41
	CFMVC-ETR (ACMMM'23)	97.39(0.00)	98.14(0.00)	96.65(0.00)	94.78(0.00)	96.68(0.00)	97.90(0.00)	6119.14
	Orth-NTF (ICCV'24)	51.26(0.00)	57.81(0.00)	46.04(0.00)	59.94(0.00)	39.84(0.00)	55.40(0.00)	363.98
	SLR-MVC	**99.38(0.00)**	**99.35(0.00)**	**99.41(0.00)**	**98.48(0.00)**	**99.21(0.00)**	**99.60(0.00)**	780.51
AwA_fea	BMVC (TPAMI'18)	5.96(0.00)	4.96(0.00)	7.47(0.00)	12.89(0.00)	3.25(0.00)	9.98(0.00)	10.21
	AWMVC (AAAI'23)	4.44(0.06)	4.74(0.07)	4.17(0.05)	10.34(0.15)	2.30(0.06)	9.09(0.12)	249.99
	FPMVS-CAG (TIP'22)	5.12(0.10)	4.76(0.07)	5.54(0.16)	11.13(0.13)	2.68(0.09)	8.55(0.13)	645.32
	S^2MVTC (CVPR'24)	59.32(0.00)	53.44(0.00)	66.66(0.00)	81.93(0.00)	58.24(0.00)	66.96(0.00)	10.80
	EBMGC-GNF (TPAMI'24)	4.56(0.00)	2.34(0.00)	**99.67(0.00)**	0.31(0.00)	-0.00(0.00)	3.84(0.00)	2579.66
	sMERA-MVC (TMM'24)	84.67(1.63)	**87.59(1.64)**	81.96(1.91)	91.56(0.46)	84.32(1.67)	86.23(1.79)	344.93
	CFMVC-ETR (ACMMM'23)	73.06(0.00)	77.02(0.00)	69.49(0.00)	89.18(0.00)	72.45(0.00)	76.70(0.00)	511.94
	Orth-NTF (ICCV'24)	OM	OM	OM	OM	OM	OM	OM
	SLR-MVC	**88.01(2.70)**	86.54(4.23)	89.61(1.80)	**94.00(0.61)**	**87.72(2.77)**	**88.62(2.64)**	158.13
CIFAR100	BMVC (TPAMI'18)	89.57(0.00)	81.76(0.00)	99.03(0.00)	98.42(0.00)	89.46(0.00)	91.27(0.00)	42.22
	AWMVC (AAAI'23)	93.12(0.99)	89.17(1.46)	97.45(0.64)	98.69(0.20)	93.05(1.00)	91.92(1.26)	377.81
	FPMVS-CAG (TIP'22)	90.64(1.96)	85.24(3.09)	96.81(0.60)	98.27(0.35)	90.54(1.98)	89.98(1.93)	1386.99
	S^2MVTC (CVPR'24)	85.89(0.00)	78.34(0.00)	95.06(0.00)	96.64(0.00)	85.73(0.00)	86.10(0.00)	6.87
	EBMGC-GNF (TPAMI'24)	**99.92(0.00)**	**99.92(0.00)**	**99.92(0.00)**	**99.94(0.00)**	**99.92(0.00)**	**99.96(0.00)**	1404.59
	sMERA-MVC (TMM'24)	94.22(0.89)	91.26(1.40)	97.39(0.32)	98.16(0.22)	94.16(0.90)	94.77(0.96)	163.26
	CFMVC-ETR (ACMMM'23)	87.81(0.00)	87.14(0.00)	88.48(0.00)	91.60(0.00)	87.68(0.00)	92.48(0.00)	778.61
	Orth-NTF (ICCV'24)	36.86(0.00)	35.33(0.00)	38.53(0.00)	70.03(0.00)	36.20(0.00)	46.67(0.00)	12418.21
	SLR-MVC	98.36(0.54)	97.04(1.00)	99.72(0.12)	99.65(0.12)	98.34(0.55)	98.36(0.56)	128.02

Table 3: The comparison of clustering results, including mean values (standard deviation), using different methods on six multi-view datasets. ("OM" indicates out of memory, and "-" signifies that the algorithm took more than four hours.)

Clustering Performance Analysis

Table 3 presents the clustering performance of various methods across six multi-view datasets, evaluated using F-score, Precision, Recall, ARI, NMI, ACC, and CPU time. The best and second-best results for each metric are highlighted in **bold** and underlined, respectively.

Firstly, among pairwise inter-view correlation-based methods, S^{2}MVTC, which considers local smoothness within each view, achieves the best performance on the ORL, CCV, Reuters, and AWA_fea datasets. It implies that these datasets exhibit strong local consistency within views. Furthermore, tensor-based MVC methods, such as sMERA-MVC, CFMVC-ETR, and SLR-MVC, generally outperform pairwise inter-view methods like BMVC, AWMVC, and FPMVS-CAG, particularly on datasets such as ORL, CCV, and Reuters. This shows the benefits of leveraging high-order inter-view correlations across different views, leading to significantly improved clustering performance.

Specifically, compared to CFMVC-ETR, which focuses on enhanced tensor rank and tensorial exclusive regularization to effectively capture inter-view high-order correlations, SLR-MVTC achieves higher ACC performance, with improvements of 2.90%, 3.88%, 11.16%, 11.29%, 1.70%, and 5.88% on the ORL, CCV, ALOI_100, Reuters, AWA_fea, and CIFAR100 datasets, respectively. This suggests the significant role of intra-view correlations. When compared to sMERA-MVC, which treats the learning of both inter- and intra-view correlations equally, SLR-MVTC demonstrates a higher F-score performance with increases of 5.44%, 5.68%, 9.93%, 9.13%, 3.34%, and 4.14% on the same datasets. This indicates that local similarities within views and global high-order correlations across views should be weighted differently. Additionally, while EBMGC-GNF achieves the best performance on CIFAR-100, our method performs similarly but is approximately ten times faster in terms of CPU time.

In conclusion, the proposed SLR-MVTC method not only outperforms state-of-the-art methods in clustering performance but also demonstrates significantly better computational efficiency.

Model Discussion

Convergence Analysis Fig. 4 illustrates the convergence performance of the proposed SLR-MVTC across six multi-view datasets, where our method converges when RE $\leq 10^{-5}$. It can be observed that RE decreases rapidly during the initial iterations on all datasets, stabilizing after approximately 10 iterations, and eventually approaches zero. Similarly, the ACC increases quickly in the first few iterations and shows minimal changes after 10 iterations. This suggests that within the first 10 iterations, both intra-view local similarity and inter-view global high-order correlations are effectively captured from the multi-view data, enhancing clustering performance. Moreover, the method converges quickly, especially on ORL, CCV, ALOI_100, and Reuters, typically reaching convergence around the 20th iteration.

Ablation Study The newly defined norm $\|\mathcal{U}\|_{\text{STNN}}$, which calculates the low-rank (LR) approximation of the low-frequency (LF) components, is crucial in the SLR-MVTC model. To further investigate its impact, we conducted two modifications: 1) replacing $\|\mathcal{U}\|_{\text{STNN}}$ with $\|\mathcal{U}\|_{\text{TNN}}$ (resulting in the "w/o LF" model), and 2) replacing $\|\mathcal{U}\|_{\text{STNN}}$ with $\|\mathcal{U}\|_{\text{TLFC}}$ (resulting in the "w/o LR" model). The results, shown in Table 4, reveal that on the CCV, ALOI-100, Reuters, and AWA-fea datasets, removing the LR component led to a slight decline in clustering performance, while removing the LF component significantly degraded performance. This suggests that inter-view and intra-view correlations have different impacts on clustering, with local smoothness exploration within views being more effective in enhancing clustering outcomes. Furthermore, compared to the S^{2}MVTC model in Table 3, which also considers local smoothness within views, "w/o LR" performs better. This improvement is likely due to S^{2}MVTC's two-stage learning process, whereas "w/o LR" directly utilizes learn latent features from multi-view data, making it more efficient.

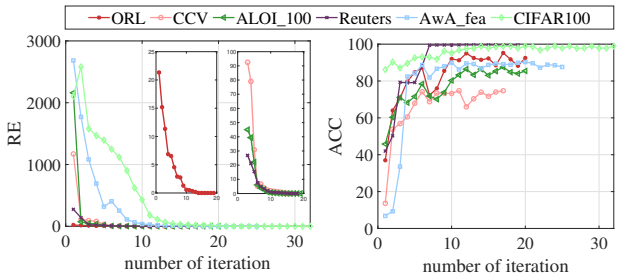


Figure 4: The convergence performance of SLR-MVTC.

Datasets	w/o LF	w/o LR	SLR-MVTC
ORL	90.80(2.58)	91.12(1.20)	**92.90(1.49)**
CCV	43.92(0.24)	72.53(3.70)	**72.84(2.07)**
ALOI_100	67.95(2.05)	84.16(1.05)	**85.49(1.51)**
Reuters	43.90(0.02)	96.44(0.00)	**99.60(0.00)**
AwA_fea	74.40(1.69)	88.49(1.13)	**88.62(2.64)**
CIFAR100	89.09(1.45)	91.25(2.67)	**98.36(0.56)**

Table 4: The accuracy of ablation study on six datasets.

Conclusion

In this paper, we focus on directly learning the shared latent features from multi-view data for clustering, where the learned features exhibit local smoothness within views and global low-rank structure across views. By minimizing the STNN norm, which optimizes the low-rank approximation of the low-frequency component on the feature tensor, SLR-MVTC effectively captures local similarities within views and global high-order correlations across views. Numerical experiments on six multi-view datasets demonstrate that SLR-MVTC outperforms all state-of-the-art methods.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (NSFC) (No.6240011126,

No.62020106011, No.62171088), the Sichuan Science and Technology Program (No.2024NSFSC1473), the Central Guidance for Local Science and Technology Development Fund Project (No.2024ZYD0268) and the China Postdoctoral Science Foundation (No.2024M750356).

References

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; and Eckstein, J. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1): 1–122.

Braman, K. 2010. Third-order tensors as linear operators on a space of matrices. *Linear Algebra and its Applications*, 433(7): 1241–1253.

Cui, C.; Ren, Y.; Pu, J.; Pu, X.; and He, L. 2023. Deep multi-view subspace clustering with anchor graph. In *IJCAI*, IJCAI '23. ISBN 978-1-956792-03-4.

Geusebroek, J.-M.; Burghouts, G. J.; and Smeulders, A. W. 2005. The Amsterdam library of object images. *IJCV*, 61: 103–112.

He, Z.; Wan, S.; Zappatore, M.; and Lu, H. 2023. A similarity matrix low-rank approximation and inconsistency separation fusion approach for multiview clustering. *TAI*, 5(2): 868–881.

Huang, Z.; Ren, Y.; Pu, X.; and He, L. 2021. Non-linear fusion for self-paced multi-view clustering. In *ACMMM*, 3211–3219.

Huang, Z.; Ren, Y.; Pu, X.; Huang, S.; Xu, Z.; and He, L. 2023. Self-supervised graph attention networks for deep weighted multi-view clustering. In *AAAI*, volume 37, 7936–7943.

Huizing, G.-J.; Deutschmann, I. M.; Peyré, G.; and Cantini, L. 2023. Paired single-cell multi-omics data integration with Mowgli. *Nature Communications*, 14(1): 7711.

Ji, J.; and Feng, S. 2023. High-order complementarity induced fast multi-view clustering with enhanced tensor rank minimization. In *ACMMM*, 328–336.

Jiang, Y.; Ye, G.; Chang, S.; Ellis, D.; and Loui, A. 2011. Consumer video understanding: A benchmark database and an evaluation of human and machine performance. In *Proc. ACM Con. Multimedia Retrieval*, 1–8.

Khan, A.; and Maji, P. 2019. Approximate graph Laplacians for multimodal data clustering. *TPAMI*, 43(3): 798–813.

Kilmer, M.; Braman, K.; Hao, N.; and Hoover, R. 2013. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.*, 34(1): 148–172.

Kilmer, M. E.; and Martin, C. D. 2011. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3): 641–658.

Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.

Lampert, C. H.; Nickisch, H.; and Harmeling, S. 2009. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 951–958. IEEE.

Lewis, D. D.; Yang, Y.; Russell-Rose, T.; and Li, F. 2004. Rcv1: A new benchmark collection for text categorization research. *JMLR*, 5(Apr): 361–397.

Li, J.; Gao, Q.; WANG, Q.; Yang, M.; and Xia, W. 2023. Orthogonal Non-negative Tensor Factorization based Multi-view Clustering. In *NeurIPS*, volume 36, 18186–18202.

Liu, J.; Liu, X.; Yang, Y.; Liu, L.; Wang, S.; Liang, W.; and Shi, J. 2021. One-pass multi-view clustering for large-scale data. In *ICCV*, 12344–12353.

Liu, J.; Wang, C.; Gao, J.; and Han, J. 2013. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM international conference on data mining*, 252–260. SIAM.

Liu, Y.; Chen, J.; Lu, Y.; Ou, W.; Long, Z.; and Zhu, C. 2024. Adaptively topological tensor network for multi-view subspace clustering. *TKDE*.

Liu, Y.; He, L.; Cao, B.; Yu, P.; Ragin, A.; and Leow, A. 2018. Multi-view multi-graph embedding for brain network clustering analysis. In *AAAI*, volume 32.

Long, Z.; Wang, Q.; Ren, Y.; Liu, Y.; and Zhu, C. 2024. S2MVTC: a Simple yet Efficient Scalable Multi-View Tensor Clustering. In *CVPR*, 26213–26222.

Long, Z.; Zhu, C.; Chen, J.; Li, Z.; Ren, Y.; and Liu, Y. 2023. Multi-view MERA Subspace Clustering. *TMM*, 1–11.

Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2020. Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm. *TPAMI*, 42(4): 925–938.

Lu, Y.; Liu, Y.; Long, Z.; Chen, Z.; and Zhu, C. 2023. O-minus decomposition for multi-view tensor subspace clustering. *TAI*.

Qian, X.; Pei, J.; Zheng, H.; Xie, X.; Yan, L.; Zhang, H.; Han, C.; Gao, X.; Zhang, H.; Zheng, W.; et al. 2021. Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning. *Nature biomedical engineering*, 5(6): 522–532.

Ren, Y.; Pu, J.; Yang, Z.; Xu, J.; Li, G.; Pu, X.; Philip, S. Y.; and He, L. 2024. Deep clustering: A comprehensive survey. *TNNLS*.

Samaria, F. S.; and Harter, A. C. 1994. Parameterisation of a stochastic model for human face identification. In *Proceedings of 1994 IEEE workshop on applications of computer vision*, 138–142. IEEE.

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

Wang, J.; Tian, F.; Yu, H.; Liu, C. H.; Zhan, K.; and Wang, X. 2017. Diverse non-negative matrix factorization for multiview data representation. *TCYB*, 48(9): 2620–2632.

Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2022. Fast Parameter-Free Multi-View Subspace Clustering With Consensus Anchor Guidance. *TIP*, 31: 556–568.

Wu, D.; Yang, Z.; Lu, J.; Xu, J.; Xu, X.; and Nie, F. 2024. EBMGC-GNF: Efficient Balanced Multi-view Graph Clustering via Good Neighbor Fusion. *TPAMI*.

Xia, W.; Gao, Q.; Wang, Q.; Gao, X.; Ding, C.; and Tao, D. 2022. Tensorized bipartite graph learning for multi-view clustering. *TPAMI*.

Xie, D.; Gao, Q.; Deng, S.; Yang, X.; and Gao, X. 2021. Multiple graphs learning with a new weighted tensor nuclear norm. *Neural Networks*, 133: 57–68.

Xie, Y.; Tao, D.; Zhang, W.; Liu, Y.; Zhang, L.; and Qu, Y. 2018. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *IJCV*, 126(11): 1157–1179.

Xu, C.; Si, J.; Guan, Z.; Zhao, W.; Wu, Y.; and Gao, X. 2024. Reliable conflictive multi-view learning. In *AAAI*, volume 38, 16129–16137.

Xu, C.; Zhao, W.; Zhao, J.; Guan, Z.; Yang, Y.; Chen, L.; and Song, X. 2023a. Progressive deep multi-view comprehensive representation learning. In *AAAI*, volume 37, 10557–10565.

Xu, J.; Ren, Y.; Tang, H.; Yang, Z.; Pan, L.; Yang, Y.; Pu, X.; Yu, P. S.; and He, L. 2023b. Self-Supervised Discriminative Feature Learning for Deep Multi-View Clustering. *TKDE*, 35(7): 7470–7482.

Yan, W.; Zhang, Y.; Lv, C.; Tang, C.; Yue, G.; Liao, L.; and Lin, W. 2023. Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In *CVPR*, 19863–19872.

Yu, X.; Xu, M.; Zhang, Y.; Liu, H.; Ye, C.; Wu, Y.; Yan, Z.; Zhu, C.; Xiong, Z.; Liang, T.; et al. 2023. Mvimgnet: A large-scale dataset of multi-view images. In *CVPR*, 9150–9161.

Zhang, C.; Jia, X.; Li, Z.; Chen, C.; and Li, H. 2024. Learning Cluster-Wise Anchors for Multi-View Clustering. In *AAAI*, volume 38, 16696–16704.

Zhang, R.; Nie, F.; Li, X.; and Wei, X. 2019. Feature selection with multi-view data: A survey. *Information Fusion*, 50: 158–167.

Zhang, Z.; Liu, L.; Shen, F.; Shen, H. T.; and Shao, L. 2018. Binary multi-view clustering. *TPAMI*, 41(7): 1774–1782.

Zhou, T.; Ruan, S.; and Canu, S. 2019. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array*, 3: 100004.