

# Unified View Extraction with Low-Rankness and Smoothness Fusion for Multi-View Subspace Clustering

Yapeng Wang<sup>1</sup>, Quanxue Gao<sup>2\*</sup>, Fangfang Li<sup>2</sup>, Yu Yun<sup>2</sup>, Ming Yang<sup>3\*</sup>

<sup>1</sup> School of Mathematics and Statistics, Xidian University, Shaanxi 710071, China

<sup>2</sup> School of Telecommunications Engineering, Xidian University, Shaanxi 710071, China

<sup>3</sup> College of Mathematical Sciences, Harbin Engineering University, Heilongjiang 150001, China  
ypwang0121@stu.xidian.edu.cn, qxgao@xidian.edu.cn, 22011110201@stu.xidian.edu.cn, 569650354@qq.com, yangmingmath@gmail.com

## Abstract

Tensor-based multi-view subspace clustering (MVSC) has achieved significant success by capturing high-order inter-view correlations. However, existing approaches face two principal limitations. First, most methods either exclusively emphasize the inter-view **low-rankness** (R) prior while neglecting the intra-view **local smoothness** (S) prior, or treat R and S as two separate regularizers—complicating joint optimization. Second, conventional **tensor-based** methods impose only low-rank constraints on the representation tensor, which limits their ability to simultaneously model consistency and complementary information. To address these issues, we propose a *Unified View Extraction with Low-Rankness and Smoothness Fusion (UVELRS)* method. Our framework first extracts a consistent cross-view representation and then constructs a tensor by stacking these representations. We introduce a novel *tensor total variation Schatten- $p$  norm* that simultaneously encodes both R and S priors while offering flexible singular-value control. This unified formulation effectively captures both high-order inter-view correlations and intra-view local smoothness. Extensive experiments on real-world datasets demonstrate UVELRS’s superior performance and robustness.

**Code** — <https://github.com/xd-wyp/Appendix>

## Introduction

Multi-view data, ubiquitous in real-world applications, consists of datasets collected from diverse sources or described by heterogeneous feature dimensions pertaining to the same objects. For instance, image quality assessment requires integrating complementary features such as sharpness, noise characteristics, and structural integrity. Multi-view subspace clustering (MVSC) leverages complementary information across views and underlying inter-view consensus to classify unlabeled samples into categories with high intra-cluster similarity and low inter-cluster correlation, thereby revealing latent data patterns. Owing to its information-fusion capability, MVSC has achieved significant success in domains including image clustering (Jiang et al. 2025), recommender systems (Xu et al. 2025), and industrial fault diagnosis (Liu et al. 2023).

\*Corresponding authors.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Early MVSC methodologies typically employ nonnegative matrix factorization (NMF) (Wang, Zhang, and Gao 2018) or the low-rank representation (LRR) (Chen et al. 2021) to learn view-specific self-representation matrices that capture intrinsic subspace structures. These matrices are subsequently fused through consensus-matrix enforcement or joint reconstruction-error minimization to reveal the cluster membership. Recent approaches incorporate tensor representations to model high-order correlations. Conventionally, view-specific representation matrices are stacked into a third-order tensor, and the tensor nuclear norm (TNN) (Kilmer et al. 2013) is optimized to enforce low-rank structure and achieve cross-view fusion (Chen, Xiao, and Zhou 2020; Li et al. 2023b). Recognizing the TNN’s suboptimality as a rank surrogate, emerging studies (Zhang, Guo, and Pan 2024; Xie et al. 2024) apply nonconvex mappings to singular values, achieving tighter rank approximation. These approaches aggressively suppresses negligible singular values while preserving dominant components, thereby enhancing discriminative power.

Despite promising results, TNN-based methods exhibit significant limitations. First, they frequently overlook the inherent local smoothness prior: samples within the same cluster should exhibit similar representation coefficients. Current implementations (Xie, Gao, and Yang 2023; Sun and Zhang 2024) impose this via auxiliary regularizers, which complicates joint optimization with low-rank constraints and increases model complexity. Second, sole reliance on tensor low-rank constraints inadequately captures both view consistency and complementary information, resulting in suboptimal representations. Consequently, developing a unified regularizer that integrates low-rankness and local smoothness remains crucial.

Motivated by these observations, we propose a unified tensor-based multi-view subspace clustering framework integrating low-rankness and local smoothness (UVELRS). Specifically, we first learn view-specific self-representation matrices. Departing from prior work, we derive a unified consensus representation that preserves consistency information and combine it with the view-specific matrices into a third-order tensor. Inspired by variational image processing (He et al. 2015; Wang et al. 2023), we introduce a novel tensor total variation Schatten- $p$  norm. Optimizing this norm simultaneously enforces: (1) low-rankness across views to

capture high-order correlations, and (2) local smoothness within each view. This synergistic integration of  $\mathbf{R}$  and  $\mathbf{S}$  priors significantly enhances clustering performance.

Our principal contributions are:

- A novel tensor total variation Schatten- $p$  norm unifying  $\mathbf{R}$  and  $\mathbf{S}$  priors in a single regularizer. The tunable Schatten- $p$  constraint provides fine-grained singular value control, effectively reducing redundancy while preserving essential structures.
- We propose a unified representation framework extracting a consensus view from self-representation matrices and tensorizing it with view-specific representations, ensuring inter-view consistency while capturing high-order correlations and intra-view smoothness.
- An efficient optimization algorithm based on the augmented Lagrangian multiplier method, accompanied by rigorous convergence analysis to a KKT point. Extensive experiments on six benchmarks demonstrating superior clustering accuracy and robustness versus state-of-the-art baselines.

## Notations and Preliminaries

### Notations

We adopt the following conventions: calligraphic letters denote tensor (e.g.,  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  for a third-order tensor), with  $\mathcal{A}^{(v)} \in \mathbb{R}^{n_1 \times n_2}$  representing its  $v$ -th frontal slice matrix. The identity tensor  $\mathcal{I}$  satisfies  $\mathcal{I}^{(1)} = \mathbf{I}$  and  $\mathcal{I}^{(k)} = \mathbf{0}$  for  $k > 1$ . The Fast Fourier Transform (FFT) along the third dimension is  $\overline{\mathcal{A}} = \text{fft}(\mathcal{A}, [], 3)$ , with inverse  $\mathcal{A} = \text{ifft}(\overline{\mathcal{A}}, [], 3)$ . The mode- $k$  unfolding  $\mathcal{A}_{(k)}$  satisfies  $\mathcal{A}_{(k)} \in \mathbb{R}^{n_k \times (\prod_{i \neq k} n_i)}$ , and  $\text{fold}_k(\cdot)$  reconstructs the tensor  $\text{fold}_k(\mathcal{A}_{(k)}) = \mathcal{A}$ .

### Preliminaries

**Definition 1. Mode- $k$  product** (Kilmer and Martin 2011) For  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  and matrix  $\mathbf{U} \in \mathbb{R}^{n_k \times n_k}$  ( $k \in \{1, 2, 3\}$ ), the mode- $k$  product  $\mathcal{Y} := \mathcal{A} \times_k \mathbf{U}$  satisfies

$$\mathcal{Y}_{(k)} = \mathbf{U} \mathcal{A}_{(k)}. \quad (1)$$

**Definition 2.  $t$ -product** (Kilmer and Martin 2011) For  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  and  $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ , the  $t$ -product is defined as

$$\mathcal{A} * \mathcal{B} := \text{ifft}(\overline{\mathcal{A}} \Delta \overline{\mathcal{B}}, [], 3), \quad (2)$$

where  $\Delta$  denotes the face-wise product:  $(\overline{\mathcal{A}} \Delta \overline{\mathcal{B}})^{(i)} = \overline{\mathcal{A}}^{(i)} \overline{\mathcal{B}}^{(i)}$  for each frontal slice  $i$ .

**Definition 3.  $f$ -diagonal tensor** (Kilmer and Martin 2011) A tensor is called  $f$ -diagonal when every frontal slice is a diagonal matrix.

**Definition 4. Gradient tensor** (Wang et al. 2023) For  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , the gradient tensor along the  $k$ -th mode is defined as

$$\mathcal{G}_k := \nabla_k(\mathcal{A}) = \mathcal{A} \times_k \mathbf{D}_{n_k} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \quad k = 1, 2, 3, \quad (3)$$

where  $\mathbf{D}_{n_k}$  is the circulant matrix generated by the vector  $[-1, 1, 0, \dots, 0]^T \in \mathbb{R}^{n_k}$ .

**Definition 5.  $t$ -SVD** (Kilmer and Martin 2011; Kilmer et al. 2013) The tensor SVD ( $t$ -SVD) of  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is given by

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \quad (4)$$

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 satisfying  $\mathcal{U}^T * \mathcal{U} = \mathcal{I}$  and  $\mathcal{V} * \mathcal{V}^T = \mathcal{I}$ , and  $\mathcal{S}$  is an  $f$ -diagonal tensor.

**Definition 6. Tensor Schatten- $p$  norm** (Gao et al. 2020b) The Schatten- $p$  norm of  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is defined as

$$\|\mathcal{A}\|_{\mathfrak{S}} := \left( \sum_{i=1}^{n_3} \|\overline{\mathcal{A}}^{(i)}\|_{\mathfrak{S}}^p \right)^{\frac{1}{p}} = \left( \sum_{i=1}^{n_3} \sum_{j=1}^h \sigma_j^p(\overline{\mathcal{A}}^{(i)}) \right)^{\frac{1}{p}}, \quad (5)$$

where  $p \in (0, 1]$ ,  $h = \min(n_1, n_2)$  and  $\sigma_j(\cdot)$  denotes the  $j$ -th singular value. This norm generalizes the tensor nuclear norm (TNN) when  $p = 1$  (Kilmer and Martin 2011).

## Related Work

Tensor-based multi-view subspace clustering optimizes representation matrices by capturing high-order correlations across views, achieving notable success in machine learning and pattern recognition (Cheng, Jing, and Ng 2018; Li et al. 2023a; Wu, Feng, and Yuan 2024). Given multi-view data  $\{\mathbf{X}^v\}_{v=1}^V$  where  $\mathbf{X}^v \in \mathbb{R}^{d_v \times N}$  denotes the  $v$ -th view ( $d_v$  features,  $N$  samples), a representative model is:

$$\begin{aligned} & \min_{\mathcal{Z}, \mathbf{E}} \mathcal{R}(\mathcal{Z}) + \lambda \|\mathbf{E}\|_{2,1} \\ & \text{s.t. } \mathbf{X}^v = \mathbf{X}^v \mathbf{Z}^v + \mathbf{E}^v, \quad \forall v \in \{1, \dots, V\}, \\ & \mathcal{Z} = \Phi(\mathbf{Z}^1, \dots, \mathbf{Z}^V), \quad \mathbf{E} = [\mathbf{E}^1; \dots; \mathbf{E}^V]. \end{aligned} \quad (6)$$

Here  $\mathbf{Z}^v \in \mathbb{R}^{N \times N}$  is the  $v$ -th self-representation matrix,  $\Phi(\cdot)$  stacks  $\{\mathbf{Z}^v\}_{v=1}^V$  along the third dimension to form tensor  $\mathcal{Z} \in \mathbb{R}^{N \times N \times V}$ ,  $\mathcal{R}(\mathcal{Z})$  is a regularization term for clustering enhancement, and  $\mathbf{E}$  aggregates view-specific noise matrices  $\mathbf{E}^v$ .

Recent advances in tensor theory have established tensor singular value decomposition (t-SVD) as a powerful framework, which has achieved success in tensor completion (Wang et al. 2024) and tensor robust principal component analysis (TRPCA) (Lin et al. 2024). Many MVSC methods adopt the tensor nuclear norm (TNN)  $\|\mathcal{Z}\|_*$  for  $\mathcal{R}(\mathcal{Z})$  to capture high-order cross-view correlations, outperforming matrix-based approaches. To mitigate noise sensitivity, (Gao et al. 2020a) proposed adaptive singular-value weighting, while (Pan, Li, and Che 2024) and (Ding et al. 2025) employed nonconvex regularizers to map singular values to  $[0, 1]$  for improved rank approximation. These methods typically neglect the *intra-view smoothness prior*, which compromises within-view consistency and clustering performance.

Complementary approaches encode smoothness priors through view-specific regularizations such as sliced sparse gradient (SSG) (Sun et al. 2022) or total variation (TV) (Du and Lu 2025). Such regularizations constrain representation coefficients within individual views to reduce noise and enhance intra-view coherence. However, they operate

independently per view without cross-view interaction, preventing simultaneous enforcement of smoothness across the multi-view representation. Consequently, they fail to fully leverage cross-view correlations or maintain consistent smoothing patterns, limiting the effectiveness of multi-view information fusion.

Thus, existing approaches either ignore the  $\mathbf{S}$ -prior or isolate the  $\mathbf{S}$ - and  $\mathbf{R}$ -priors in separate regularization terms, leading to suboptimal joint optimization. This results in incomplete utilization of hidden structures in multi-view data.

## The Proposed Method

This section introduces a unified view representation  $\widehat{\mathbf{Z}}$  reconstructed from view-specific matrices  $\{\mathbf{Z}^v\}_{v=1}^V$  through tensorization, enhancing consistency modeling. We propose a novel *tensor total variation Schatten- $p$  norm* regularization that integrates  $\mathbf{R}$  (inter-view correlations) and  $\mathbf{S}$  (intra-view smoothness) priors, replacing conventional TNN. This design simultaneously captures local smoothness within views and high-order correlations across views.

### Model Formulation

Given multi-view data  $\{\mathbf{X}^v\}_{v=1}^V$ , we extract a unified representation  $\widehat{\mathbf{Z}}$  from  $\mathbf{Z}^v$  via:

$$\mathbf{Z}^v = \widehat{\mathbf{Z}} + \mathbf{E}_2^v. \quad (7)$$

This decomposition preserves cross-view consistency. We then construct a third-order tensor  $\mathcal{Z} \in \mathbb{R}^{N \times N \times (V+1)}$  by stacking  $\{\mathbf{Z}^v\}_{v=1}^V$  and  $\widehat{\mathbf{Z}}$ , and impose the tensor total variation Schatten- $p$  norm, yielding:

$$\begin{aligned} \min_{\mathcal{Z}, \mathbf{E}^v} \quad & \lambda_1 \|\mathcal{Z}\|_{\mathbb{S}, \text{TTV}}^p + \lambda_2 \sum_{v=1}^V \|\mathbf{E}^v\|_{2,1} \\ \text{s.t.} \quad & \mathbf{X}^v = \mathbf{X}^v \mathbf{Z}^v + \mathbf{E}_1^v, \\ & \mathbf{Z}^v = \widehat{\mathbf{Z}} + \mathbf{E}_2^v, \\ & \mathcal{Z} = \Phi(\mathbf{Z}^1, \dots, \mathbf{Z}^V, \widehat{\mathbf{Z}}), \\ & \mathbf{E}^v = [\mathbf{E}_1^v; \mathbf{E}_2^v]. \end{aligned} \quad (8)$$

The norm  $\|\mathcal{Z}\|_{\mathbb{S}, \text{TTV}}^p$  extends variational image processing techniques (Wang et al. 2023), preserving both low-rank structure and local smoothness of  $\mathcal{Z}$ . Its Schatten- $p$  constraint provides fine-grained singular value penalization, enabling tighter rank approximation and enhanced noise suppression. We formalize this norm and justify its necessity below.

**Definition 7. Tensor Total Variation Schatten- $p$  Norm** For a tensor  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , let  $\Gamma$  be a prior mode set. The norm is:

$$\|\mathcal{Z}\|_{\mathbb{S}, \text{TTV}}^p := \sum_k \|\mathcal{G}_k\|_{\mathbb{S}}^p, \quad (9)$$

where  $\mathcal{G}_k = \nabla_k(\mathcal{A}) = \mathcal{A} \times_k \mathbf{D}_{n_k}$  represents the gradient tensor of  $\mathcal{A}$  along with the  $k$ -mode.

We provide a clarification of Definition 7. Real-world clustering tasks often involve numerous samples but few classes, where intra-class representations exhibit inherent similarity. Here,  $\Gamma$  denotes modes enforcing local smoothness. For  $\mathcal{Z}$ , we set  $\Gamma = \{1, 2\}$  because coefficients corresponding to identical samples should vary smoothly along modes 1 and 2. Consequently,  $\nabla_1(\mathcal{Z})$  and  $\nabla_2(\mathcal{Z})$  become sparse. To incorporate low-rank priors, we replace conventional  $\ell_1/\ell_2$  matrix norms with tensor-SVD and Schatten- $p$  norms, capturing high-order inter-view correlations. This approach fuses  $\mathbf{R}$  and  $\mathbf{S}$  priors into a single regularizer, avoiding multi-term regularization trade-offs, simplifying optimization, and reducing hyperparameter complexity. The Schatten- $p$  constraint further enables precise singular value control for enhanced complementary information extraction.

We perform a numerical experiment to validate the above analysis. Using the MSRCV dataset (Winn and Jovic 2005), we apply the classical t-SVD-MVC method (Xie et al. 2018) to each view to obtain its self-representation matrix, and stack these matrices into a third-order tensor  $\mathcal{Z}$ . We then compute the singular value (SV) decay curves and element-wise frequency histograms for the gradient tensors  $\nabla_1(\mathcal{Z})$  and  $\nabla_2(\mathcal{Z})$ , shown in Fig. 1. From these results, we conclude: (1) The rapid singular value decay indicates a pronounced low-rank structure, enabling effective extraction of high-order correlations across views; (2) The predominance of zero-valued elements demonstrates strong intra-view smoothness, reflecting highly similar representation coefficients for data points within the same class.

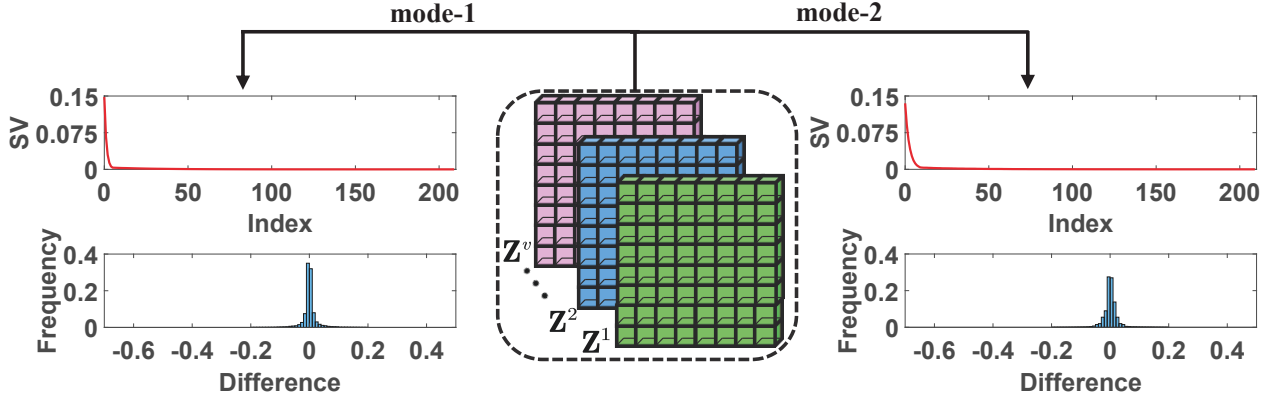
Thus, the proposed regularizer effectively fuses low-rankness and local smoothness structures, enabling unified multi-view modeling while enhancing clustering performance and representation robustness.

### Optimization

To solve Eq. (8) with multiple variables, we employ the Alternating Direction Method of Multipliers (ADMM), updating one variable per iteration while fixing others. Introducing two auxiliary variables  $\mathcal{A} = \mathcal{Z}$  and  $\mathcal{G}_k = \nabla_k(\mathcal{A})$  yields the augmented Lagrangian function:

$$\begin{aligned} \mathcal{L}(\mathcal{Z}, \mathcal{A}, \mathcal{G}_k, \mathbf{E}^v) = & \lambda_1 \sum_k \|\mathcal{G}_k\|_{\mathbb{S}}^p + \lambda_2 \sum_{v=1}^V \|\mathbf{E}^v\|_{2,1} \\ & + \sum_{v=1}^V \left( \langle \mathbf{J}^v, \mathbf{X}^v - \mathbf{X}^v \mathbf{Z}^v - \mathbf{E}_1^v \rangle + \frac{\mu}{2} \|\mathbf{Z}^v - \widehat{\mathbf{Z}} - \mathbf{E}_2^v\|_{\text{F}}^2 \right) \\ & + \sum_{v=1}^V \left( \langle \mathbf{O}^v, \mathbf{Z}^v - \widehat{\mathbf{Z}} - \mathbf{E}_2^v \rangle + \frac{\mu}{2} \|\mathbf{X}^v - \mathbf{X}^v \mathbf{Z}^v - \mathbf{E}_1^v\|_{\text{F}}^2 \right) \\ & + \sum_k \left( \langle \mathcal{N}^k, \mathcal{G}_k - \nabla_k(\mathcal{A}) \rangle + \frac{\theta}{2} \|\mathcal{G}_k - \nabla_k(\mathcal{A})\|_{\text{F}}^2 \right) \\ & + \langle \mathcal{C}, \mathcal{Z} - \mathcal{A} \rangle + \frac{\theta}{2} \|\mathcal{Z} - \mathcal{A}\|_{\text{F}}^2, \end{aligned} \quad (10)$$

where  $\mathbf{J}^v, \mathbf{O}^v, \mathcal{N}^k, \mathcal{C}$  are Lagrange multipliers, and  $\mu, \theta > 0$  denote penalty parameters. We solve these subproblems iteratively:



(a) Low-rankness and smoothness of  $\nabla_1(\mathcal{Z})$  (b) Self-presentation tensor  $\mathcal{Z}$  (c) Low-rankness and smoothness of  $\nabla_2(\mathcal{Z})$

Figure 1: Illustration of low-rankness and local smoothness in gradient tensors  $\nabla_1(\mathcal{Z})$  and  $\nabla_2(\mathcal{Z})$ .

(1) **Update  $\mathbf{Z}^v$ :** With other variables fixed:

$$\begin{aligned} \min_{\mathbf{Z}^v} \frac{\mu}{2} \left\| \mathbf{Z}^v - \widehat{\mathbf{Z}} - \mathbf{E}_2^v + \frac{\mathbf{O}^v}{\mu} \right\|_F^2 + \frac{\theta}{2} \left\| \mathbf{Z}^v - \mathbf{A}^v + \frac{\mathbf{C}^v}{\theta} \right\|_F^2 \\ + \frac{\mu}{2} \left\| \mathbf{X}^v - \mathbf{X}^v \mathbf{Z}^v - \mathbf{E}_1^v + \frac{\mathbf{J}^v}{\mu} \right\|_F^2. \end{aligned} \quad (11)$$

Setting the derivative of Eq. (11) to zero gives:

$$\mathbf{Z}^v = (\mathbf{K}_1^v)^{-1} \mathbf{K}_2^v, \quad (12)$$

where  $\mathbf{K}_1^v = (\mu + \theta)\mathbf{I} + \mu(\mathbf{X}^v)^\top \mathbf{X}^v$  and  $\mathbf{K}_2^v = \mu(\mathbf{X}^v)^\top (\mathbf{X}^v - \mathbf{E}_1^v + \mu^{-1} \mathbf{J}^v) + \mu (\widehat{\mathbf{Z}} + \mathbf{E}_2^v - \mu^{-1} \mathbf{O}^v) + \theta (\mathbf{A}^v - \theta^{-1} \mathbf{C}^v)$ .

(2) **Update  $\mathbf{E}^v$ :** With other variables fixed:

$$\min_{\mathbf{E}^v} \frac{\lambda_2}{\mu} \|\mathbf{E}^v\|_{2,1} + \frac{1}{2} \|\mathbf{E}^v - \mathbf{G}^v\|_F^2, \quad (13)$$

where  $\mathbf{G}^v = [\mathbf{G}_1^v; \mathbf{G}_2^v]$  with  $\mathbf{G}_1^v = \mathbf{X}^v - \mathbf{X}^v \mathbf{Z}^v + \mu^{-1} \mathbf{J}^v$  and  $\mathbf{G}_2^v = \mathbf{Z}^v - \widehat{\mathbf{Z}} + \mu^{-1} \mathbf{O}^v$ . Following (Liu et al. 2012), the closed-form solution of Eq. (13) is:

$$\mathbf{E}^v(:, i) = \begin{cases} \left(1 - \frac{\lambda_2/\mu}{\|\mathbf{G}^v(:, i)\|_2}\right) \mathbf{G}^v(:, i), & \|\mathbf{G}^v(:, i)\|_2 > \frac{\lambda_2}{\mu} \\ \mathbf{0}, & \text{otherwise.} \end{cases} \quad (14)$$

(3) **Update  $\widehat{\mathbf{Z}}$ :** With other variables fixed:

$$\begin{aligned} \min_{\widehat{\mathbf{Z}}} \sum_{v=1}^V \frac{\mu}{2} \|\mathbf{Z}^v - \widehat{\mathbf{Z}} - \mathbf{E}_2^v\|_F^2 \\ + \frac{\theta}{2} \left\| \widehat{\mathbf{Z}} - \mathbf{A}^{(V+1)} + \frac{\mathbf{C}^{(V+1)}}{\theta} \right\|_F^2. \end{aligned} \quad (15)$$

Setting the derivative of Eq. (15) to zero yields:

$$\widehat{\mathbf{Z}} = \frac{\mathbf{K}_3}{\theta + \mu V}, \quad (16)$$

where  $\mathbf{K}_3 = \sum_{v=1}^V (\mathbf{O}^v + \mu(\mathbf{Z}^v - \mathbf{E}_2^v)) - \mathbf{C}^{(V+1)} + \theta \mathbf{A}^{(V+1)}$ .

(4) **Update  $\mathcal{G}_k$ :** With other variables fixed:

$$\min_{\mathcal{G}_k} \lambda_1 \|\mathcal{G}_k\|_{\mathbb{S}}^p + \frac{\theta}{2} \left\| \nabla_k(\mathcal{A}) - \frac{\mathcal{N}^k}{\theta} - \mathcal{G}_k \right\|_F^2. \quad (17)$$

Using Lemma 1 from (Gao et al. 2020b):

**Lemma 1.** For  $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  with  $t$ -SVD:  $\mathcal{S} = \mathcal{U} * \Sigma * \mathcal{V}^\top$ , the optimal solution of

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{X} - \mathcal{S}\|_F^2 + \tau \|\mathcal{X}\|_{\mathbb{S}}^p \quad (18)$$

is  $\mathcal{X}^* = \Gamma_\tau(\Sigma) = \mathcal{U} * \text{ifft}(\mathbf{C}_\tau(\overline{\Sigma})) * \mathcal{V}^\top$ , where  $\mathbf{C}_\tau$  denotes the generalized shrinkage thresholding (Gao et al. 2020b).

Thus, the solution to Eq. (17) is:

$$\mathcal{G}_k = \Gamma_{\frac{\lambda_1}{\theta}} \left( \nabla_k(\mathcal{A}) - \frac{\mathcal{N}^k}{\theta} \right). \quad (19)$$

(5) **Update  $\mathcal{A}$ :** With other variables fixed:

$$\min_{\mathcal{A}} \sum_k \left\| \mathcal{G}_k - \nabla_k(\mathcal{A}) + \frac{\mathcal{N}^k}{\theta} \right\|_F^2 + \left\| \mathcal{Z} - \mathcal{A} + \frac{\mathcal{C}}{\theta} \right\|_F^2. \quad (20)$$

Setting the derivative to zero gives:

$$\left( \mathcal{I} + \sum_k \nabla_k^\top \nabla_k \right) \mathcal{A} = \mathcal{Z} + \frac{\mathcal{C}}{\theta} + \sum_k \nabla_k^\top \left( \mathcal{G}_k + \frac{\mathcal{N}^k}{\theta} \right), \quad (21)$$

where  $\nabla_k^\top$  is the transpose operator of  $\nabla_k$ . Following (Wang et al. 2023), we solve Eq. (21) in the Fourier domain:

$$\mathcal{A} = \mathcal{F}^{-1} \left( \frac{\mathcal{H} + \mathcal{F}(\mathcal{Z} + \frac{\mathcal{C}}{\theta})}{\mathbf{1} + \sum_k \mathcal{F}(\mathcal{D}_k)^\top \odot \mathcal{F}(\mathcal{D}_k)} \right), \quad (22)$$

where  $\mathcal{H} = \sum_k \mathcal{F}(\mathcal{D}_k)^\top \odot \mathcal{F} \left( \mathcal{G}_k + \frac{\mathcal{N}^k}{\theta} \right)$ ,  $\mathbf{1}$  denotes an all-ones tensor,  $\odot$  represents component-wise multiplication, and division is component-wise.

## (6) Update Multipliers and Penalty Parameters:

$$\begin{aligned}
\mathbf{J}^v &= \mathbf{J}^v + \mu (\mathbf{X}^v - \mathbf{X}^v \mathbf{Z}^v - \mathbf{E}_1^v); \\
\mathbf{O}^v &= \mathbf{O}^v + \mu (\mathbf{Z}^v - \hat{\mathbf{Z}} - \mathbf{E}_2^v); \\
\mathcal{N}^k &= \mathcal{N}^k + \theta (\mathcal{G}_k - \nabla_k(\mathcal{A})); \\
\mathcal{C} &= \mathcal{C} + \theta (\mathcal{Z} - \mathcal{A}); \\
\mu &= \min(\rho\mu, \mu_{\max}); \quad \theta = \min(\rho\theta, \theta_{\max}).
\end{aligned} \tag{23}$$

The complete optimization procedure is outlined in Algorithm 1.

## Convergence Analysis

**Theorem 1.** (The Convergence Analysis of Algorithm 1) Let  $P_t = \{\mathcal{Z}_t, \mathbf{E}_{1,t}^v, \mathbf{E}_{2,t}^v, \mathbf{J}_t^v, \mathbf{O}_t^v, \mathcal{N}_t^k, \mathcal{G}_{k,t}, \mathcal{C}_t, \mathcal{A}_t\}$  denotes the iteration sequence generated during the optimization process in Algorithm 1. Then we have

- (1)  $\{P_t\}$  is bounded;
- (2) Any accumulation point of Algorithm 1 is a KKT point.

**Corollary 1** (Geometric Convergence). Assume multipliers are bounded:  $\|\mathbf{J}^v\|_F \leq M_J$ ,  $\|\mathbf{O}^v\|_F \leq M_O$ ,  $\|\mathcal{N}^k\|_F \leq M_N$ ,  $\|\mathcal{C}\|_F \leq M_C$ . With adaptive penalties  $\mu_t = \mu_0 \rho^t$ ,  $\theta_t = \theta_0 \rho^t$  ( $\rho > 1$ ), the primal residual decays linearly:

$$R_t \leq C \beta^t,$$

where  $C = \max\left(\frac{2\rho M_J}{\mu_0}, \frac{2\rho M_O}{\mu_0}, \frac{2\rho M_N}{\theta_0}, \frac{2\rho M_C}{\theta_0}\right)$ ,  $\beta = \rho^{-1}$ . This implies  $\mathcal{O}(\ln \epsilon^{-1})$  iteration complexity.

We provide a rigorous mathematical proof of Theorem 1 in the Appendix. Under general conditions, the algorithm converges to a point and attains an exponential convergence rate. Moreover, the KKT conditions derived in the proof can also serve as practical stopping criteria for Algorithm 1, ensuring that the algorithm terminates when a stationary point is reached. In practice, the algorithm is considered to have converged once the total relative error  $\text{RE} \leq 10^{-5}$ , where  $\text{RE} = \max\left(\frac{1}{V} \sum_{v=1}^V \|\mathbf{X}^v - \mathbf{X}^v \mathbf{Z}^v - \mathbf{E}_1^v\|_\infty, \frac{1}{V} \sum_{v=1}^V \|\mathbf{Z}^v - \hat{\mathbf{Z}} - \mathbf{E}_2^v\|_\infty, \|\mathcal{Z} - \mathcal{A}\|_\infty\right)$ .

## Complexity Analysis

The overall computational cost is dominated by subproblems (11)–(20).

- $\mathbf{Z}^v$  involves matrix inversion, costing  $\mathcal{O}(VN^3)$ .
- $\mathbf{E}^v$  requires computing  $\mathbf{E}_1^v$  and  $\mathbf{E}_2^v$ , costing  $\mathcal{O}(2VN^2)$ .
- $\hat{\mathbf{Z}}$  costs  $\mathcal{O}(N^2)$ .
- Tensor subproblems (19) and (20) involve FFT and t-SVD operations, costing  $\mathcal{O}(VN^2 \log N)$  and  $\mathcal{O}(VN^2 \log N + V^2N^2)$  respectively.

The total complexity is  $\mathcal{O}\left(VN^3 + (V+1)^2N^2 + 2VN^2 \log N\right)$ .

## Algorithm 1: Solving Problem 8

**Input:** Multi-view data  $\{\mathbf{X}^v\}_{v=1}^V \in \mathbb{R}^{d_v \times N}$ , clusters  $K$ , parameters  $\lambda_1, \lambda_2$ , prior set  $\Gamma = \{1, 2\}$ .

**Initialize:** Set all matrix variables as zero matrices, and all tensor variables as zero tensors,  $\mu = \theta = 1e-3$ ,  $\mu_{\max} = \theta_{\max} = 1e10$ ,  $\rho = 1.5$ .

**Output:**  $\mathbf{Z}^v, \hat{\mathbf{Z}}$

- 1: **while** not converge **do**
- 2:   Update  $\mathbf{Z}^v$  by Eq. (12);
- 3:   Update  $\mathbf{E}^v$  by Eq. (14);
- 4:   Update  $\hat{\mathbf{Z}}$  by Eq. (16);
- 5:   Update  $\mathcal{G}_k$  by Eq. (19);
- 6:   Update  $\mathcal{A}$  by Eq. (22);
- 7:   Update  $\mathbf{J}^v, \mathbf{O}^v, \mathcal{N}^k, \mathcal{C}$  by Eq. (23);
- 8: **end while**
- 9: Using matrix  $S = \left(\frac{|\hat{\mathbf{Z}}| + |\hat{\mathbf{Z}}|^\top}{2} + \frac{1}{V} \sum_{v=1}^V \frac{|\mathbf{Z}^v| + |\mathbf{Z}^v|^\top}{2}\right)$  for spectral clustering.

## Experiments

This section presents a series of comparative experiments between our method and several state-of-the-art MVC methods on six real-world benchmark datasets. We employ three widely recognized clustering performance metrics to evaluate each MVC method, namely Accuracy (ACC), Normalized Mutual Information (NMI), and Purity (PUR). Each method is executed five times, and results are reported as the mean  $\pm$  standard deviation. In general, higher values of these metrics indicate better clustering performance.

## Datasets and Baselines

Dataset	$N$	$V$	$K$	Size
NGs	500	3	5	(2000,2000,2000)
BBCSport	544	2	5	(3183,3203)
HW	2000	6	10	(240,76,216,47,64,6)
Scene15	4485	3	15	(1800,1180,1240)
MITindoor-67	5360	4	67	(3600,1770,1240,4096)
CCV	6773	3	20	(20,20,20)

Table 1: The multi-view datasets in our experiments.

These six widely used datasets are NGs, BBCSport (Winn and Jojic 2005), HW (Dua, Graff et al. 2017), Scene15 (Oliva and Torralba 2001), MITIndoor (Quattoni and Torralba 2009), and CCV (Jiang et al. 2011). The detailed information of these datasets is presented in Table 1.

We compare the proposed method with the following MVC approaches: (1) Classical Tensor-based MVC (t-SVD-MVC) (Xie et al. 2018); (2) Weighted Tensor-Nuclear Norm Minimization (WTNNM) (Gao et al. 2020a); (3) Sliced Sparse Gradient Norm Minimization with Enhanced Tensor Rank (SSG-TAR) (Sun et al. 2022); (4) Clean Non-Convex Tensorized MVSC (NMSC-MCP) (Zhang, Guo, and Pan 2024); (5) Feature Projection with Automatic Dimensionality Selection (TLRADS) (Cai et al. 2025); (6) Low-

### Experimental Results

Table 2 details the clustering performance of all methods across six datasets using ACC, NMI, and PUR metrics, with best results highlighted in bold. These results demonstrate that our proposed UVELRS consistently outperforms all baseline approaches. Specifically, we draw the following conclusions:

(1) Unlike t-SVD-MVC, WTNNM, and NMSC-MVC which impose only low-rank constraints on the representation tensor, UVELRS additionally incorporates an intra-view  $\mathbf{S}$ -prior. By enforcing smooth variation of representation coefficients among same-class samples within each view, our method ensures highly consistent representations for similar samples. This approach effectively reduces noise and eliminates spurious connections. The combined global low-rank constraint and local-smooth structures preserve both inter-view correlations and intra-view consistency, yielding more accurate and robust clustering across diverse datasets.

(2) Compared to SSG-TAR and LS-LRTR that apply separate smoothing regularization to each view-specific matrix, UVELRS directly enforces local smoothness on the unified representation tensor. This tensor-level formulation integrates low-rankness and local smoothness into a single regularization term, enabling coordinated modeling of inter-view correlations and intra-view coherence. Unlike methods handling these priors separately, UVELRS eliminates manual tuning of multiple regularization weights, simplifies the parameter space, and enhances optimization stability.

### Impact of Parameters

We analyze the influence of parameters  $\lambda_1$ ,  $\lambda_2$ , and  $p$  on clustering performance using the NGs and Scene15 datasets.

**Impact of  $\lambda_1$  and  $\lambda_2$**  For all datasets, we evaluate  $\lambda_1, \lambda_2 \in \{0.01, 0.05, 0.1, 1, 5, 10, 100\}$ . Figure 2 shows their impact on clustering accuracy (ACC) for NGs and Scene15 datasets. The ACC metric remains consistently high when these parameters are stable. While fixing  $\lambda_1$  and varying  $\lambda_2$  (or vice versa) induces minor ACC fluctuations, the results indicate considerable parameter robustness. This mutual compensation effect suggests appropriate but non-critical parameter selection in practical applications.

**Impact of  $p$**  We varied the parameter  $p$  from 0.1 to 1 in increments of 0.1. This produced the clustering performance on the NGs and Scene15 datasets, measured by ACC, NMI, and PUR, as shown in Figure 3. Therefore, we can conclude that the value of  $p$  has a significant impact on clustering performance. This is mainly because the Schatten- $p$  constraint, as a non-convex regularization, provides a better approximation to the rank function, thereby enabling more effective extraction of complementary information across views.

### Experiment of Convergence

We also empirically evaluated the convergence behavior of our method. Specifically, we plotted the reconstruction er-

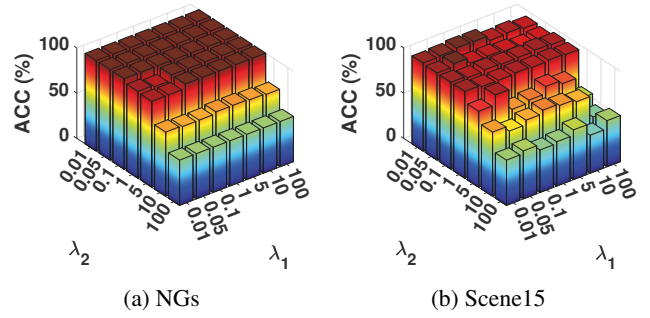


Figure 2: Sensitivity analysis of parameters  $\lambda_1$  and  $\lambda_2$  on Clustering ACC for (a) NGs and (b) Scene15.

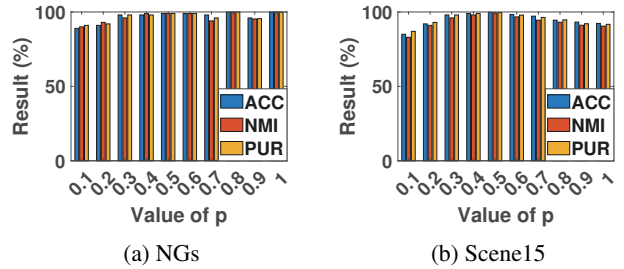


Figure 3: Clustering performance under different values of  $p$  on (a) NGs and (b) Scene15.

rors:  $\text{Error}_1 = \frac{1}{V} \sum_{v=1}^V \|\mathbf{X}^v - \mathbf{X}^v \mathbf{Z}^v - \mathbf{E}_1^v\|_\infty$ ,  $\text{Error}_2 = \frac{1}{V} \sum_{v=1}^V \|\mathbf{Z}^v - \hat{\mathbf{Z}} - \mathbf{E}_2^v\|_\infty$  and the matching error:  $\text{Error}_3 = \|\mathbf{Z} - \hat{\mathbf{A}}\|_\infty$  over iterations on the NGs and Scene15 datasets, as shown in Figure 4. From the curves, we observe that all three errors rapidly decrease within the first 10 iterations and then stabilize at low values, demonstrating the efficiency and stability of our optimization scheme.

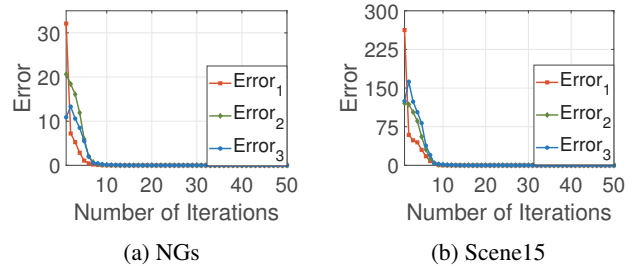


Figure 4: The convergence curves on (a) NGs and (b) Scene15.

### t-SNE Visualization of Clustering Results

Figures 5 and 6 illustrate the t-SNE visualizations for the HW and BBCSport datasets, respectively. In the initial iterations, the sample points are largely intermingled, resulting in poor cluster separation. As the algorithm converges, the t-SNE plots reveal clearly defined clusters: samples belonging to the same class become tightly grouped, while differ-

Dataset	Metrics	t-SVD-MVC	WTNNM	SSG-TAR	NMSC-MCP	TLRADS	LS-LRTR	UVELRS
NGs	ACC	92.33(0.17)	94.93(0)	96.42(0.13)	97.21(0.17)	95.24(0.38)	97.60(0.41)	<b>100(0)</b>
	NMI	94.51(0.66)	96.42(0)	97.33(0.52)	98.99(0.36)	98.77(0.16)	98.21(0.39)	<b>100(0)</b>
	PUR	93.78(0.54)	94.38(0)	96.98(0.12)	97.34(0.22)	97.43(0.27)	99.13(0.11)	<b>100(0)</b>
BBCSport	ACC	94.53(0.17)	96.42(0)	96.58(0.04)	99.43(0.13)	94.63(0.26)	97.63(0.02)	<b>100(0)</b>
	NMI	95.82(0.25)	97.92(0)	97.23(0.16)	99.56(0.07)	93.45(0.38)	98.51(0.06)	<b>100(0)</b>
	PUR	95.77(0.71)	97.68(0)	97.44(0.13)	99.28(0.19)	92.19(0.22)	98.74(0.13)	<b>100(0)</b>
HW	ACC	92.98(0.16)	94.15(0)	93.44(0.23)	91.27(0.46)	93.41(0.17)	97.62(0.38)	<b>99.73(0.02)</b>
	NMI	90.37(0.22)	95.48(0)	90.26(0.17)	94.58(0.27)	94.39(0.33)	95.24(0.25)	<b>99.90(0.08)</b>
	PUR	93.13(0.01)	93.77(0)	92.75(0.24)	85.17(0.05)	90.21(0.21)	97.44(0.67)	<b>99.68(0.05)</b>
Scene15	ACC	67.32(0.58)	90.41(0)	90.91(0.55)	86.82(0.35)	91.05(0.26)	91.43(0.04)	<b>99.78(0.04)</b>
	NMI	71.55(0.39)	93.18(0)	90.76(0.38)	89.41(0.17)	89.37(0.11)	92.53(0.06)	<b>99.32(0.12)</b>
	PUR	69.36(0.77)	93.39(0)	92.43(0.71)	92.50(0.36)	90.43(0.16)	94.87(0.11)	<b>99.64(0.09)</b>
MITIndoor-67	ACC	68.45(0.17)	76.34(0.77)	94.51(0)	79.51(0.38)	88.26(0.17)	94.42(0.31)	<b>99.43(0.27)</b>
	NMI	75.03(0.26)	81.69(0.46)	98.73(0)	91.43(0.88)	86.17(0.05)	96.66(0.64)	<b>99.57(0.18)</b>
	PUR	71.44(0.57)	80.01(0.58)	96.44(0)	94.55(0.16)	85.18(0.22)	95.77(0.51)	<b>99.44(0.36)</b>
CCV	ACC	42.95(0.06)	47.46(0.23)	59.47(0.13)	57.78(0.69)	54.37(0.28)	59.32(0.27)	<b>76.43(0.47)</b>
	NMI	39.86(0.05)	50.21(0.12)	52.55(0.51)	61.22(0.16)	60.24(0.15)	61.11(0.54)	<b>72.85(0.34)</b>
	PUR	32.57(0.04)	46.83(0.33)	59.57(0.14)	59.41(0.34)	57.85(0.37)	60.77(0.73)	<b>75.91(0.16)</b>

Table 2: The clustering performance of all methods on six datasets, with each metric reported as the mean (standard deviation).

ent classes progressively separate. This visual progression closely mirrors the steady increase in clustering accuracy observed over successive iterations.

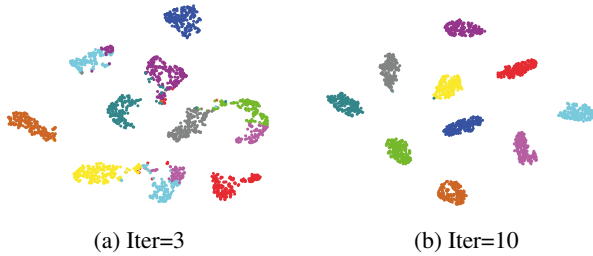


Figure 5: t-SNE visualization results on HW with different numbers of iterations.

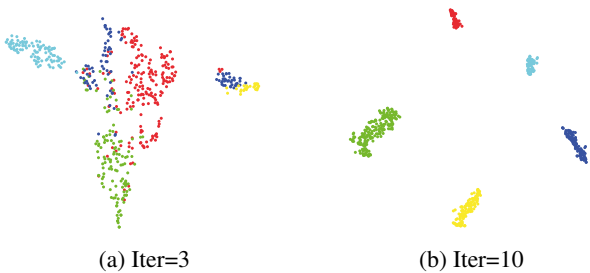


Figure 6: t-SNE visualization results on BBCSport with different numbers of iterations.

### Ablation Study

An ablation study was conducted by removing the unified view extraction  $\hat{\mathbf{Z}}$  (denoted as w.o.  $\hat{\mathbf{Z}}$ ) and replacing the tensor total variation Schatten- $p$  norm  $\|\mathcal{Z}\|_{\mathbb{D},\text{TTV}}^p$  with the tensor Schatten- $p$  norm  $\|\mathcal{Z}\|_{\mathbb{D}}^p$  (denoted as “only  $\|\mathcal{Z}\|_{\mathbb{D}}^p$ ”). The

optimal value of parameter  $p$  in these models is determined through experiments. As shown in Table 3, the removal of  $\hat{\mathbf{Z}}$  led to a modest decline in clustering accuracy, indicating that enforcing inter-view consistency is beneficial. In contrast, substituting  $\|\mathcal{Z}\|_{\mathbb{D},\text{TTV}}^p$  with the  $\|\mathcal{Z}\|_{\mathbb{D}}^p$  resulted in a substantial drop in performance, as the model lost its ability to simultaneously capture low-rankness and local smoothness. These findings emphasize the importance of both components for achieving optimal clustering performance.

Dataset	only $\ \mathcal{Z}\ _{\mathbb{D}}^p$	w.o. $\hat{\mathbf{Z}}$	UVELRS
NGs	94.47(0.11)	96.11(0.25)	100(0)
BBCSport	89.63(0.18)	97.14(0.33)	100(0)
HW	87.32(0.55)	95.74(0.19)	99.73(0.02)
SCene15	73.56(0.18)	95.33(0.77)	99.78(0.04)
MITIndoor-67	73.66(0.48)	94.65(0.55)	99.43(0.27)
CCV	44.42(0.33)	64.38(0.51)	76.43(0.47)

Table 3: The result of ablation study.

### Conclusion

In this work, we presented a unified MVSC model that jointly incorporates tensor low-rankness and local smoothness priors. The proposed approach first constructs a unified representation to preserve consistency across views, which is then fused into a tensor structure. To fully exploit both high-order correlations among views and local smoothness within each view, we introduced a novel regularization term that effectively integrates these two forms of prior knowledge. An efficient optimization scheme was devised to solve the proposed model, with guaranteed convergence to a KKT point. Through extensive experiments on six widely used benchmark datasets, the proposed method consistently outperformed several state-of-the-art baselines, validating its robustness and strong clustering capability.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China, Grant No. 62176203 and 62576263; the Natural Science Basic Research Program of Shaanxi Province, Grant No. 2025JC-QYCX-051; the Fundamental Research Funds for the Central Universities and the Innovation Fund of Xidian University, Grant No. YJSJ25007.

## References

- Cai, B.; Lu, G.-F.; Guo, X.; and Wu, T. 2025. Tensorized latent representation with automatic dimensionality selection for multi-view clustering. *Pattern Recognition*, 160: 111192.
- Chen, J.; Yang, S.; Mao, H.; and Fahy, C. 2021. Multiview subspace clustering using low-rank representation. *IEEE Transactions on Cybernetics*, 52(11): 12364–12378.
- Chen, Y.; Xiao, X.; and Zhou, Y. 2020. Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix. *Pattern Recognition*, 106: 107441.
- Cheng, M.; Jing, L.; and Ng, M. K. 2018. Tensor-based low-dimensional representation learning for multi-view clustering. *IEEE Transactions on Image Processing*, 28(5): 2399–2414.
- Ding, M.; Yang, J.-H.; Zhao, X.-L.; Zhang, J.; and Ng, M. K. 2025. Nonconvex Low-Rank Tensor Representation for Multi-View Subspace Clustering With Insufficient Observed Samples. *IEEE Transactions on Knowledge and Data Engineering*.
- Du, Y.; and Lu, G.-F. 2025. Joint local smoothness and low-rank tensor representation for robust multi-view clustering. *Pattern Recognition*, 157: 110944.
- Dua, D.; Graff, C.; et al. 2017. UCI machine learning repository.
- Gao, Q.; Xia, W.; Wan, Z.; Xie, D.; and Zhang, P. 2020a. Tensor-SVD based graph learning for multi-view subspace clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 3930–3937.
- Gao, Q.; Zhang, P.; Xia, W.; Xie, D.; Gao, X.; and Tao, D. 2020b. Enhanced tensor RPCA and its application. *IEEE transactions on pattern analysis and machine intelligence*, 43(6): 2133–2140.
- He, W.; Zhang, H.; Zhang, L.; and Shen, H. 2015. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE transactions on geoscience and remote sensing*, 54(1): 178–188.
- Jiang, M.; Hu, L.; He, Z.; and Chen, Z. 2025. Interpretable multi-view clustering. *Pattern Recognition*, 111418.
- Jiang, Y.-G.; Ye, G.; Chang, S.-F.; Ellis, D.; and Loui, A. C. 2011. Consumer video understanding: A benchmark database and an evaluation of human and machine performance. In *Proceedings of the 1st ACM international conference on multimedia retrieval*, 1–8.
- Kilmer, M. E.; Braman, K.; Hao, N.; and Hoover, R. C. 2013. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications*, 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.
- Li, J.; Gao, Q.; Wang, Q.; Yang, M.; and Xia, W. 2023a. Orthogonal non-negative tensor factorization based multi-view clustering. *Advances in neural information processing systems*, 36: 18186–18202.
- Li, X.; Sun, Y.; Sun, Q.; Ren, Z.; and Sun, Y. 2023b. Cross-view graph matching guided anchor alignment for incomplete multi-view clustering. *Information Fusion*, 100: 101941.
- Lin, J.; Huang, T.-Z.; Zhao, X.-L.; Ji, T.-Y.; and Zhao, Q. 2024. Tensor robust kernel PCA for multidimensional data. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2): 2662–2674.
- Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2012. Robust recovery of subspace structures by low-rank representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(1): 171–184.
- Liu, X.; Wang, J.; Meng, S.; Qiu, X.; and Zhao, G. 2023. Multi-view rotating machinery fault diagnosis with adaptive co-attention fusion network. *Engineering Applications of Artificial Intelligence*, 122: 106138.
- Oliva, A.; and Torralba, A. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42: 145–175.
- Pan, B.; Li, C.; and Che, H. 2024. Error-robust multi-view subspace clustering with nonconvex low-rank tensor approximation and hyper-Laplacian graph embedding. *Engineering Applications of Artificial Intelligence*, 133: 108274.
- Quattoni, A.; and Torralba, A. 2009. Recognizing indoor scenes. In *2009 IEEE conference on computer vision and pattern recognition*, 413–420. IEEE.
- Sun, X.; Zhu, R.; Yang, M.; Zhang, X.; and Tang, Y. 2022. Sliced sparse gradient induced multi-view subspace clustering via tensorial arctangent rank minimization. *IEEE Transactions on Knowledge and Data Engineering*, 35(7): 7483–7496.
- Sun, Y.; and Zhang, F. 2024. Low-rank multi-view subspace clustering based on sparse regularization. *Journal of Computer and Communications*, 12(4): 14–30.
- Wang, H.; Peng, J.; Qin, W.; Wang, J.; and Meng, D. 2023. Guaranteed tensor recovery fused low-rankness and smoothness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9): 10990–11007.
- Wang, P.; Su, Y.; Huang, B.; Zhu, D.; Liu, W.; Nedzved, A.; Krasnoproschin, V. V.; and Leung, H. 2024. Low rank tensor completion pansharpening based on haze correction. *IEEE Transactions on Geoscience and Remote Sensing*.
- Wang, X.; Zhang, T.; and Gao, X. 2018. Multiview clustering based on non-negative matrix factorization and pairwise measurements. *IEEE transactions on cybernetics*, 49(9): 3333–3346.

- Winn, J.; and Jojic, N. 2005. Locus: Learning object classes with unsupervised segmentation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, 756–763. IEEE.
- Wu, T.; Feng, S.; and Yuan, J. 2024. Low-rank kernel tensor learning for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 15952–15960.
- Xie, D.; Gao, Q.; and Yang, M. 2023. Enhanced tensor low-rank representation learning for multi-view clustering. *Neural Networks*, 161: 93–104.
- Xie, D.; Yang, M.; Gao, Q.; and Song, W. 2024. Non-convex tensorial multi-view clustering by integrating  $l_1$ -based sliced-Laplacian regularization and  $l_2, p$ -sparsity. *Pattern Recognition*, 154: 110605.
- 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. *International Journal of Computer Vision*, 126: 1157–1179.
- Xu, Z.; Chen, S.; Pan, W.; and Ming, Z. 2025. A multi-view graph contrastive learning framework for cross-domain sequential recommendation. *ACM Transactions on Recommender Systems*, 3(4): 1–28.
- Zhang, X.; Guo, X.; and Pan, J. 2024. A new nonconvex multi-view subspace clustering via learning a clean low-rank representation tensor. *Inverse Problems*, 40(12): 125007.