BEYOND SVD: DIFFERENTIABLE LOW-RANK MATRIX FACTORIZATION FOR MULTIVIEW SUBSPACE CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-view subspace clustering (MvSC) has attracted considerable attention due to its ability to uncover underlying subspace structures from heterogeneous multi-view data. However, existing methods require time-consuming SVD during each iteration, resulting in quadratic or even cubic computational complexity with respect to the number of samples and thereby limiting their scalability to large-scale datasets. To overcome these limitations, we propose a novel differentiable low-rank matrix factorization method for MvSC, which unifies multi-factor decomposition, low-rank representation learning, and structural regularization within a coherent framework. The proposed method not only characterizes the intrinsic low-rank structure through multiple factor matrices under cross-view consistency and symmetry constraints, but also transforms the factor-wise iterative optimization into differentiable learning steps, avoiding repeated SVD and enabling end-to-end learning through back propagation. The resulting block-wise optimization algorithm is further supported by a theoretical convergence analysis. Extensive experiments on multiple benchmark datasets demonstrate that our method achieves superior clustering performance while substantially reducing computational costs.

1 INTRODUCTION

Multiview data from diverse sources present substantial analytical challenges due to high dimensionality and heterogeneous structures Fu et al. (2026b); Liu et al. (2026). To this end, multiview subspace clustering (MvSC) has emerged as a promising paradigm for learning representations that reveal latent subspace structures across views. In recent years, MvSC has attracted considerable attention in the machine learning and computer vision communities and has been widely applied to various tasks such as face clustering Xie et al. (2018); Tang et al. (2023), image segmentation Liu et al. (2025); Luo et al. (2025), and anomaly detection Hou et al. (2020); Wang et al. (2026).

Rank minimization is central to MvSC methods based on low rank representation, where the nuclear norm is commonly adopted as a tractable surrogate for the rank function Xie et al. (2018); Shan et al. (2023). However, nuclear norm regularization tends to excessively shrink dominant singular values, introducing bias into the recovered representations Wang et al. (2024b); Ding et al. (2025); Wang et al. (2025b). To alleviate this limitation, nonconvex surrogates have been incorporated into MvSC to impose magnitude dependent penalties on singular values and better preserve dominant components Wang et al. (2020); Gu & Feng (2024). Despite their representational advantages, these methods require repeated, time-consuming singular value decomposition (SVD), incurring substantial computational costs as the number of samples increases Liu et al. (2024); Qin & Feng (2026). To alleviate this computational bottleneck, MvSC frameworks utilizing anchors and factorization have emerged as highly effective alternatives in recent years Wan et al. (2023); Chen et al. (2026c). Anchor-based methods reduce the storage and computational costs of large coefficient matrices by replacing sample-to-sample relationships with more compact sample-to-anchor associations Qin et al. (2025); Chen et al. (2026b). However, their clustering performance depends critically on the quality and representativeness of the selected anchors. Alternatively, factorization approaches decompose the representation matrix or tensor into a product of lower dimensional factors. While this reduces computational overhead, enforcing explicit spectral regularization on these factors often still necessitates SVD during the update procedure Wang et al. (2023); Yan et al. (2026).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Alongside efforts to improve computational efficiency, deep neural networks have been increasingly integrated into MvSC frameworks to learn more expressive representations. Representative methods integrate Tensor Transformer autoencoders with adaptive contrastive learning guided by self expression Wang et al. (2025a), incorporate pseudo label guidance into deep contrastive self representation Yu et al. (2025), or adopt noise invariant joint optimization to improve clustering robustness Chen et al. (2026a). Although these deep models achieve promising empirical performance, they typically require extensive network training and offer limited interpretability of the optimization process. These limitations motivate the integration of optimization models with trainable networks to preserve explicit modeling assumptions while learning components of the solution process Zhang & Ghanem (2018); Yang et al. (2016); Chen et al. (2023); Zhao et al. (2026). In particular, Chen et al. (2023) developed a differentiable low rank learning framework for matrix completion that replaces costly repeated SVD with trainable transformations and optimizes their parameters through backpropagation. However, directly extending this paradigm to MvSC faces fundamental structural and optimization challenges. Unlike matrix completion, which targets recovery of individual entries within a single data space, MvSC must simultaneously ensure reconstruction fidelity within each view, enforce latent consistency across views, and incorporate geometric constraints for clustering, such as symmetry and block diagonal structure Gao et al. (2026); Zhou et al. (2026); Fu et al. (2026a). These heterogeneous constraints tightly couple factor learning across views, substantially increasing the difficulty of differentiable optimization. Consequently, a novel mechanism must be fundamentally constructed to integrate multiview factor coordination deeply with differentiable low rank factorization.

To address these challenges, we propose differentiable low rank matrix factorization (DLRMF), a framework for MvSC that jointly learns low rank factors across views through differentiable optimization. The proposed approach not only captures consistent and complementary structures across views but also avoids repeated SVD during factor updates, reducing computational cost. Compared with existing MvSC methods, the proposed method offers two complementary advantages. First, it combines explicit low rank regularization with consistency and symmetry constraints to learn structured representations for clustering. Second, it enables efficient parameter learning through backpropagation while retaining an explicit connection to the underlying optimization model. The main contributions are summarized as follows:

- We propose a novel MvSC model that simultaneously imposes low rank constraints on multiple factor matrices, promotes consistency across views, and encourages symmetry within each self representation matrix. It provides greater flexibility in approximating the intrinsic low-rank structure and yields more structured representations for clustering.
- We establish a differentiable low-rank factorization framework for MvSC by incorporating the coupled factor updates into trainable modules. This design replaces repeated SVD during factor updates with learnable low rank transformations, reducing computational cost while maintaining an explicit connection between the network operations and the optimization formulation.
- Extensive experiments across multiple benchmark datasets demonstrate that our method achieves consistent gains over representative state-of-the-art approaches in terms of both clustering performance and computational efficiency.

2 Related work

Low-Rank Representation (LRR) is a prevailing subspace clustering approach that captures global sample relationships through a learned low-rank coefficient matrix Liu et al. (2013). Based on the self-expressive assumption, each sample can be represented by a linear combination of other samples from the same view. For multi-view data $\{X^{(v)}\}_{v=1}^{V}$, a general MvSC formulation can be written as

$$\min_{Z^{(v)}\in\mathbb{C}} \sum_{v=1}^{V} \left\| X^{(v)} - X^{(v)}Z^{(v)} \right\|_l + \lambda\Omega\left(Z^{(1)},\ldots,Z^{(V)}\right). \tag{1}$$

where $Z^{(v)}$ denotes the self-representation matrix of the v-th view, $\|\cdot\|_l$ measures the reconstruction error, and $\mathbb{C}$ defines the feasible set. $\Omega(\cdot)$ couples view-specific representations to capture structural correlations across views Guo et al. (2023). Within this framework, Chen et al. (2022) introduced

nuclear norm-based low-rank representations to exploit inter-view correlations. However, the nuclear norm may lead to biased low-rank estimation; accordingly, Wang et al. (2024b) introduced nonconvex regularization for a more flexible characterization of the underlying subspace structure. However, these methods typically incur quadratic or even higher computational complexity with respect to the number of samples, their scalability becomes increasingly constrained on large-scale data. To mitigate this issue, Chen et al. (2026c) learned cluster-aware consensus anchors to construct compact sample–anchor representations, while Wan et al. (2023) employed matrix factorization to obtain low-dimensional multiview representations. Despite the reduced computational cost, iterative optimization can still involve repeated SVD operations. More recently, Du et al. (2025) incorporated deep unfolding into large-scale multiview clustering by mapping iterative updates onto dedicated network modules. Despite the delightful performance achieved by these methods, extending differentiable low-rank learning to MvSC remains challenging, as reconstruction, cross-view consistency, and clustering-specific structural constraints tightly couple factor learning across views.

3 The Proposed Method

3.1 Problem Formulation

The optimization of conventional MvSC models usually is computationally expensive, with the complexity reaching $\mathcal{O}(n^3)$, posing a major obstacle to their application to large-scale datasets. To ease this problem, we adopt matrix factorization as an effective alternative, decomposing the large representation matrix into a product of several lower-dimensional factors, thereby reducing the dominant computational complexity $\mathcal{O}(nk^2 + k^3)$, where $k \ll n$ denotes the factor dimension. Based upon this, a basic multi-factor MvSC model can be formulated as

$$\min_{\{\{Z_i^{(v)}\}_{i=1}^{I}, E^{(v)}\}_{v=1}^{V}} \sum_{v=1}^{V}\sum_{i=1}^{I}\frac{1}{p_i}\|Z_i^{(v)}\|_{S_{p_i}}^{p_i} + \lambda_1 \sum_{v=1}^{V}\|E^{(v)}\|_{2,1} + \frac{\lambda_2}{2}\sum_{v=1}^{V}\|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\|_F^2 \tag{2}$$

where $Z^{(v)} = \prod_{i=1}^{I} Z_i^{(v)}$, λ_1 and λ_2 are balance parameters. Nevertheless, Eq. (2) does not sufficiently exploit the shared and complementary structures across views, which are essential for revealing reliable sample relationships and constructing an informative affinity structure in MvSC. Hence, we introduce cross-view consistency and symmetry constraints to further strengthen the structural coherence of the learned representations. The final objective of our method is formulated as

$$\begin{aligned}\min_{\{\{Z_i^{(v)}\}_{i=1}^{I}, E^{(v)}\}_{v=1}^{V}} & \sum_{v=1}^{V}\sum_{i=1}^{I}\frac{1}{p_i}\|Z_i^{(v)}\|_{S_{p_i}}^{p_i} + \underbrace{\sum_{v=1}^{V}\|E^{(v)}\|_{2,1}}_{\mathcal{J}_{\text{err}}} + \underbrace{\sum_{v=1}^{V}\|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\|_F^2}_{\mathcal{J}_{\text{rec}}} \\ & + \frac{\theta}{2}\underbrace{\sum_{v,w=1,v\neq w}^{V}\|Z^{(v)} - Z^{(w)}\|_F^2}_{\mathcal{J}_{\text{con}}} + \frac{\tau}{2}\underbrace{\sum_{v=1}^{V}\|Z^{(v)} - (Z^{(v)})^T\|_F^2}_{\mathcal{J}_{\text{sym}}},\end{aligned} \tag{3}$$

where θ, and τ control the cross-view consistency terms and symmetry terms, respectively. Despite the efficiency of factorization, the optimization still suffers from considerable cumulative cost due to repeated SVDs for spectral regularization, particularly involving numerous factors, views or a large factor dimension k. This observation motivates us to further investigate whether the SVD-based factor updates can be replaced more efficiently. Recently, differentiable unfolding networks have emerged as a feasible mechanism for addressing this issue. Representative methods such as ISTA-Net Zhang & Ghanem (2018) and ADMM-Net Yang et al. (2016) integrate model-based priors with data-driven learning by transforming iterative optimization procedures into trainable network architectures, enabling model parameters to adapt to the underlying data characteristics. Moreover, neural network architectures can readily accommodate low-rank outputs in various learning tasks. The block-coordinate formulation of Eq. (3) naturally admits a sequence of factor-wise iterative

updates, making it well suited for algorithm unfolding. This structure motivates us to recast the conventional optimization steps as differentiable learning steps, where the involved parameters are optimized through backpropagation. The resulting design preserves explicit low-rank priors while adapting to heterogeneous views without repeated SVD.

Remark 3.1 (The benefit of multi-factor factorization) Multi-factor factorization transforms the original problem into more tractable subproblems, avoiding SVD on the full representation matrix and thereby improving efficiency for large-scale data. Moreover, the number of factors and the associated p_i values provide additional flexibility for characterizing the intrinsic low-rank structure.

Remark 3.2 (The benefit of differentiable learning steps) By unfolding the iterative solver into a finite-stage differentiable architecture, the proposed method replaces repeated SVD-based thresholding with trainable factor updates and learns parameters end-to-end through backpropagation, while preserving low-rank and multiview structural priors and the explicit correspondence between network stages and optimization steps.

3.2 Optimization

The proposed model is optimized in a block-wise manner, with the i-th factor of the v-th view updated while the remaining factors are fixed. For simplicity, the last four terms in Eq. 3 are denoted by $\mathcal{J}(Z_i^{(v)})$. The proximal gradient method is then applied to update $Z_i^{(v)}$ as

$$\min_{Z_i^{(v)}} J(Z_i^{(v)}) + \frac{1}{p_i}\|Z_i^{(v)}\|_{S_{p_i}}^{p_i}. \tag{4}$$

Following the proximal gradient method, the update of $Z_i^{(v)}$ at the t-th iteration is given by

$$\begin{aligned} Z_i^{(v,t+1)} &= \arg\min_{Z_i^{(v)}} \mathcal{J}(Z_i^{(v,t)}) + \left\langle \nabla\mathcal{J}(Z_i^{(v,t)}), Z_i^{(v)} - Z_i^{(v,t)} \right\rangle + \frac{L_i^{(v)}}{2}\|Z_i^{(v)} - Z_i^{(v,t)}\|_F^2 + \frac{1}{p_i}\|Z_i^{(v)}\|_{S_{p_i}}^{p_i} \\ &= \arg\min_{Z_i^{(v)}} \frac{L_i^{(v)}}{2}\|Z_i^{(v)} - Q_i^{(v,t)}\|_F^2 + \frac{1}{p_i}\|Z_i^{(v)}\|_{S_{p_i}}^{p_i}, \end{aligned} \tag{5}$$

where $Q_i^{(v,t)} = Z_i^{(v,t)} - \frac{1}{\rho_i^{(v)} L_i^{(v)}}\nabla J(Z_i^{(v,t)})$. Here, $L_i^{(v)}$ denotes the Lipschitz constant of $\nabla_{Z_i^{(v)}}\mathcal{J}$, is given by

$$L_i^{(v)} = \max\left\{ \left[\lambda_2 \|X^{(v)} A_i^{(v)}\|_2^2 + (2\theta(V-1) + 4\tau)\, \|A_i^{(v)}\|_2^2 \right] \|B_i^{(v)}\|_2^2, \epsilon \right\}, \tag{6}$$

where $A_i^{(v)} = \prod_{j=1}^{i-1} Z_j^{(v)}$, $B_i^{(v)} = \prod_{j=i+1}^{I} Z_j^{(v)}$. Accordingly, the update of $Z_i^{(v)}$ can be expressed in terms of the Schatten-p_i proximal operator as

$$Z_i^{(v,t+1)} = \mathrm{Prox}_{S_{p_i}}\left(Z_i^{(v,t)} - \frac{1}{L_i^{(v)}}\nabla\mathcal{J}(Z_i^{(v,t)}) \right). \tag{7}$$

However, directly solving the above proximal subproblem generally involves non-differentiable SVD computations. To obtain a differentiable alternative, we further consider a stricter constrained formulation by enforcing $Q_i^{(v)} = Z_i^{(v)}$. Specifically, the factor-wise subproblem is written as

$$\arg\min_{Q_i^{(v)}} \frac{1}{p_i}\left\|Q_i^{(v)}\right\|_{S_{p_i}}^{p_i}, \quad \text{s.t.} \quad Q_i^{(v)} = Z_i^{(v)}. \tag{8}$$

The solution of Eq. (8) can be obtained by introducing a Lagrange multiplier $\Lambda_i^{(v)}$ for the equality constraint, yielding the following Lagrangian

$$\mathcal{L}\left(Q_i^{(v)}, \Lambda_i^{(v)}\right) = \frac{1}{p_i}\mathrm{Tr}\left[\left((Q_i^{(v)})^T Q_i^{(v)}\right)^{p_i/2}\right] - \mathrm{Tr}\left[(\Lambda_i^{(v)})^T\left(Q_i^{(v)} - Z_i^{(v)}\right)\right]. \tag{9}$$

Taking the partial derivative of $\mathcal{L}\left(Q_i^{(v)}, \Lambda_i^{(v)}\right)$ with respect to $Q_i^{(v)}$ and setting it to zero, we obtain

$$\frac{\partial\mathcal{L}}{\partial Q_i^{(v)}} = 2Q_i^{(v)} \cdot \frac{1}{2}\left((Q_i^{(v)})^T Q_i^{(v)}\right)^{(p_i-2)/2} - \Lambda_i^{(v)} = 0. \tag{10}$$

Accordingly, the multiplier satisfies

$$\Lambda_i^{(v)} = 2Q_i^{(v)} \cdot \frac{1}{2}\left((Q_i^{(v)})^T Q_i^{(v)}\right)^{(p_i-2)/2}. \tag{11}$$

For brevity, we define $P_i^{(v)} = \frac{1}{2}\left((Q_i^{(v)})^T Q_i^{(v)}\right)^{(p_i-2)/2}$. Then, Eq. (11) can be simplified as $\Lambda_i^{(v)} = 2Q_i^{(v)} P_i^{(v)}$. By applying the equality constraint $Q_i^{(v)} = Z_i^{(v)}$ in Eq. (8), we further obtain $\Lambda_i^{(v)} = 2Z_i^{(v)} P_i^{(v)}$. Hence, the i-th factor of the v-th view can be expressed as

$$Z_i^{(v)} = \frac{1}{2}\Lambda_i^{(v)}\left(P_i^{(v)}\right)^{-1}. \tag{12}$$

The factor-update relation derived from the Lagrangian is parameterized as a learnable mapping and embedded into a block-wise network. Specifically, the Lagrange multiplier $\Lambda_i^{(v)}$ is replaced by a trainable weight matrix $W_i^{(v)}$. Accordingly, the i-th factor of the learned representation matrix for view v is produced through the following differentiable network layer

$$P_i^{(v)} = \frac{1}{2}\left((Q_i^{(v)})^T Q_i^{(v)}\right)^{\frac{p_i-2}{2}}, \quad Z_i^{(v)} = \frac{1}{2}W_i^{(v)}\left(P_i^{(v)}\right)^{\dagger}. \tag{13}$$

Under this parameterization, the update form obtained from the Lagrangian function is incorporated into the network, and the weight matrices are optimized by back propagation using the training loss. Motivated by accelerated proximal gradient techniques Toh & Yun (2010), we introduce a learnable parameter $\rho_i^{(v)} \in (0,1)$ for the i-th factor of view v to adaptively control the step size during training. Consequently, $Q_i^{(v)}$ at each block is computed as follows:

$$Q_i^{(v)} = Z_i^{(v)} - \frac{1}{\rho_i^{(v)} L_i^{(v)}}\nabla_{Z_i^{(v)}}\mathcal{J}, \tag{14}$$

where the gradient of $\mathcal{J}(Z_i^{(v)})$ with respect to $Z_i^{(v)}$ is given by

$$\begin{aligned}\nabla_{Z_i^{(v)}}\mathcal{J} = &\lambda_2 (A_i^{(v)})^T (X^{(v)})^T \left(X^{(v)}Z^{(v)} + E^{(v)} - X^{(v)}\right)(B_i^{(v)})^T \\ &+ 2\theta (A_i^{(v)})^T \sum_{w\neq v}\left(Z^{(v)} - Z^{(w)}\right)(B_i^{(v)})^T + 2\tau (A_i^{(v)})^T \left(Z^{(v)} - (Z^{(v)})^T\right)(B_i^{(v)})^T.\end{aligned} \tag{15}$$

After obtaining the updated representation factors, the coefficient matrix of the v-th view is reconstructed as $Z^{(v)} = \prod_{i=1}^{I} Z_i^{(v)}$.

With $Z^{(v)}$ fixed, the corresponding error matrix $E^{(v)}$ is updated by solving following subproblem

$$\min_{E^{(v)}} \lambda_1 \mathcal{J}_{err} + \frac{\lambda_2}{2}\mathcal{J}_{rec}. \tag{16}$$

The above subproblem can be solved in closed form by applying column-wise $\ell_{2,1}$-norm shrinkage Yang et al. (2009)

$$E_{:,j}^{(v)} = \max\left(1 - \frac{\lambda_1}{\lambda_2 \|R_{:,j}^{(v)}\|_2 + \epsilon}, 0\right) R_{:,j}^{(v)}, \tag{17}$$

where $R^{(v)} = X^{(v)} - X^{(v)}Z^{(v)}$ and $\epsilon > 0$ is a small constant introduced to avoid numerical instability. The error matrix $E^{(v)}$ is updated independently through the above closed-form operator and is not treated as a learnable variable in back propagation. After all factors and error matrices are updated, the loss function of the proposed model is calculated as

$$\mathcal{J} = \lambda_1 \mathcal{J}_{err} + \frac{\lambda_2}{2}\mathcal{J}_{rec} + \frac{\theta}{2}\mathcal{L}_{con} + \frac{\tau}{2}\mathcal{J}_{sym}. \tag{18}$$

During training, the learnable weight matrices $W_i^{(v)}$ and step-control parameters $\rho_i^{(v)}$ are updated through back propagation until the loss $\mathcal{J}$ converges. The resulting view-specific low-rank representations are then fused to construct an affinity matrix from the angular information of its principal directions, followed by spectral clustering for final partitioning. The training and clustering procedures are provided in Appendix.

3.3 Convergence Analysis

The proposed method employs multiple differentiable factor-update blocks to enable end-to-end learning through gradient propagation. Since convergence of the training loss does not necessarily guarantee the convergence of the Schatten-p_i norm of each factor, we further analyze its descent and convergence properties through the following lemma.

Lemma 1. Nie et al. (2012) *For any positive definite matrices $A, A_t \in \mathbb{R}^{m\times m}$, the following inequality holds for $0 < p_i \leq 2$:*

$$\mathrm{Tr}\left(A^{\frac{p_i}{2}}\right) - \frac{p_i}{2}\mathrm{Tr}\left(AA_t^{\frac{p_i-2}{2}}\right) \leq \mathrm{Tr}\left(A_t^{\frac{p_i}{2}}\right) - \frac{p_i}{2}\mathrm{Tr}\left(A_tA_t^{\frac{p_i-2}{2}}\right). \tag{19}$$

Subsequently, invoking Lemma 1, Theorem 1 guarantees that the Schatten-p norm of each factor converges under forward propagation through the neural network.

Theorem 1. [1] *For $0 < p_i \leq 2$, let $Q_i^{(v)}$ and $Z_i^{(v)}$ denote the input and output of the low-rank updating operator, respectively. Then, throughout the iterative optimization process, the following inequality holds, $\left\|Z_i^{(v)}\right\|_{S_{p_i}}^{p_i} \leq \left\|Q_i^{(v)}\right\|_{S_{p_i}}^{p_i}$.*

Computational Complexity Analysis. The computational cost of DLRMF mainly arises from factor-wise gradient computation and low-rank updates. For one epoch, updating I factors over V views requires $\mathcal{O}(IVdnk + IV^2nk^2 + IVk^3)$, while the update of $E^{(v)}$ is dominated by the factor updates. For the right boundary factor, the $n \times n$ Gram matrix is avoided by using the equivalent $k \times k$ Gram matrix, reducing the corresponding cost from $\mathcal{O}(n^3)$ to $\mathcal{O}(nk^2 + k^3)$. Therefore, over T training epochs, the overall complexity is $\mathcal{O}\big(T(IVdnk + IV^2nk^2 + IVk^3)\big)$, which further reduces to $\mathcal{O}(TIVdnr + TIV^2nr^2)$ when $r \ll \min(d, n)$.

4 Experiment

4.1 Experiment Setup

Datasets. For a comprehensive empirical study, experiments are carried out on eight representative multiview benchmark datasets, including BBCSport, 100Leaves, BDGP, Wikipedia, MNIST4, Fashion, Hdigit and NUSWIDEOBJ. Among these datasets, Fashion, Hdigit, and NUSWIDEOBJ are large-scale datasets containing more than 10,000 samples. Detailed information on the datasets can be found in the Appendix.

Baselines. To further validate the competitiveness of our method, we compare it with thirteen representative multiview clustering methods. Six of them are matrix-based methods, including MLRR Chen et al. (2022) , SMSC-AND Ma et al. (2024), NLRSC Wang et al. (2024b) , AMvSC Chen et al. (2025a), SLMVGC Tan et al. (2023) and LSGMC Lan et al. (2024). Three are tensor-based methods, including T-UMC Lin et al. (2024), TAD-MVC Chen et al. (2025b), and OLR-MVC Xiao et al. (2025). Three are anchor-based methods, including CAL Chen et al. (2026c), SMVAGC Wang et al. (2024a), and TC-MVSC Chang et al. (2024). One is a deep unfolding-based method, namely LargeMvC-Net Du et al. (2025).

Evaluation Metrics. Six widely used clustering metrics are employed for performance evaluation Manning et al. (2008), namely accuracy (ACC), normalized mutual information (NMI), adjusted Rand index (ARI), F-score, Precision, and Recall. For all these criteria, a larger value corresponds to better clustering performance. Due to space limitations, the results for the three metrics, F-score, Precision, and Recall, can be found in the Appendix. The proposed DLRMF is implemented in PyTorch, with all experiments except NUSWIDEOBJ dataset conducted on an M4 processor with 16 GB RAM and NUSWIDEOBJ dataset evaluated on a high-performance computing (HPC) cluster.

Parameter Setting. The parameters of each baseline are selected through grid search based on the settings provided in the corresponding literature, and the best results are used for comparison. The five parameters of the proposed model are tuned through grid search. Specifically, $\lambda_1 \in \{0.001, 0.01, 0.03, 0.05, 0.1, 0.5\}$, $\lambda_2 \in \{0.01, 0.1, 0.3, 0.5, 0.7, 1, 1.5\}$, $\theta \in$

[1] The proofs of all theoretical results are provided in the Appendix.

Table 1: Clustering performance comparison in terms of ACC, NMI, and ARI.

Dataset	Metrics	CAL	SMVAGC	TAD-MVC	TC-MVSC	LargeMvC-Net	MLRR	SMSC-AND	NLRSC	AMvSC	SLMVGC	LSGMC	T-UMC	OLR-MVC	Ours
100Leaves	ACC	80.00	64.77	64.13	50.31	72.44	80.26	62.12	82.25	73.75	67.25	82.19	57.51	82.25	**88.88**
	NMI	92.34	85.60	82.21	59.78	86.86	91.42	80.87	91.05	90.03	82.59	92.60	76.34	91.52	**94.14**
	ARI	74.28	49.94	53.07	5.00	63.07	73.39	49.66	77.80	65.86	56.07	82.09	41.16	75.36	**82.15**
BDGP	ACC	62.60	43.60	56.72	29.88	54.68	45.64	58.82	48.96	54.00	38.94	57.44	30.71	58.40	**64.32**
	NMI	34.33	22.10	36.84	4.54	28.64	18.47	40.34	26.71	32.73	18.25	37.74	5.42	34.31	**40.48**
	ARI	29.02	17.18	23.51	4.48	24.11	14.77	34.11	22.23	29.26	14.01	30.39	23.54	31.12	**37.79**
BBCSport	ACC	94.12	85.56	80.33	78.68	86.21	93.20	59.96	90.26	94.12	76.36	95.22	75.22	96.50	**97.06**
	NMI	83.66	75.12	59.53	60.33	73.72	83.85	49.66	82.98	83.33	54.85	85.42	64.13	88.39	**90.17**
	ARI	84.07	77.72	55.22	59.56	77.45	87.24	33.46	90.26	85.77	45.79	90.02	63.61	91.17	**92.31**
Wikipedia	ACC	35.06	55.54	30.74	51.78	51.23	57.09	37.26	57.33	18.56	52.70	17.36	53.87	51.33	**61.06**
	NMI	16.77	51.06	13.95	48.34	42.78	51.71	27.41	51.93	4.22	44.68	0.39	52.44	**53.58**	52.55
	ARI	11.25	41.93	9.60	30.95	35.79	43.90	19.45	44.62	3.11	35.74	0.11	45.04	36.00	**46.32**
MNIST	ACC	89.80	88.30	85.82	89.90	75.52	78.17	61.08	64.52	64.45	91.00	65.45	64.08	80.58	**92.23**
	NMI	74.32	70.43	68.04	75.82	54.84	51.45	52.36	64.67	66.10	77.87	64.75	62.49	69.54	**80.60**
	ARI	75.95	72.48	68.07	76.21	52.04	51.36	47.58	57.61	58.83	78.73	58.18	61.61	69.16	**81.34**
Hdigit	ACC	93.66	66.01	92.72	78.40	67.10	53.15	OT	80.01	95.17	77.95	79.87	OT	48.42	**98.85**
	NMI	85.46	59.88	83.72	72.84	61.51	37.67	OT	80.50	90.55	73.07	81.29	OT	59.05	**96.80**
	ARI	86.47	50.85	84.48	53.40	52.09	36.43	OT	73.83	89.94	65.12	74.94	OT	42.35	**97.47**
Fashion	ACC	73.55	74.14	69.05	40.67	65.20	OT	OT	78.73	67.03	53.54	79.35	OT	49.85	**80.37**
	NMI	67.97	74.56	72.61	50.59	71.50	OT	OT	71.82	66.06	59.52	76.00	OT	71.31	**77.66**
	ARI	57.92	63.19	59.80	29.17	57.37	OT	OT	46.57	53.24	42.55	67.48	OT	43.02	**68.93**
NUSWIDEOBJ	ACC	12.17	13.12	12.96	OT	14.30	OT	OT	OT	OT	15.79	11.71	OT	OT	**19.62**
	NMI	9.55	12.79	10.94	OT	12.24	OT	OT	OT	OT	**14.92**	9.29	OT	OT	11.49
	ARI	2.45	4.24	3.78	OT	4.89	OT	OT	OT	OT	4.61	2.07	OT	OT	**5.00**

Note: "OT" indicates that the method exceeds the allowed running time.

Table 2: Runtime comparison (sec.) of thirteen multiview clustering methods on eight datasets.

Dataset	CAL	SMVAGC	TAD-MVC	TC-MVSC	LargeMvC-Net	MLRR	SMSC-AND	NLRSC	AMvSC	SLMVGC	LSGMC	T-UMC	OLR-MVC	Ours
BBCSport	7.48	6.38	18.74	3.76	**1.61**	25.27	8.97	29.15	86.95	3.60	6.95	36.19	6.80	5.81
100leaves	666.50	21.53	77.76	19.66	**0.53**	235.37	231.95	142.95	195.17	17.53	70.92	851.84	45.69	6.91
BDGP	56.24	21.91	93.72	106.45	**4.00**	1210.91	1022.32	344.86	373.44	86.74	62.62	2489.28	259.62	10.07
Wikipedia	42.36	2.28	83.07	69.71	**0.63**	1448.69	1895.21	434.19	288.10	71.86	57.73	3398.60	246.54	1.33
Mnist4	28.35	6.26	104.84	128.52	**0.43**	6221.25	1340.74	3638.99	2417.36	80.70	1858.94	6130.04	936.41	5.26
Hdigit	225.97	130.07	375.03	240.57	**11.25**	45315.44	OT	32544.59	33635.23	2304.85	1358.71	OT	11407.58	38.82
Fashion	361.15	220.07	182.15	398.58	**19.25**	OT	OT	61426.93	62354.44	1696.3	3293.17	OT	15509.59	62.44
NUSWIDEOBJ	2972.55	404.96	2303.07	OT	**37.74**	OT	OT	OT	OT	15351.15	9707.39	OT	OT	425.67

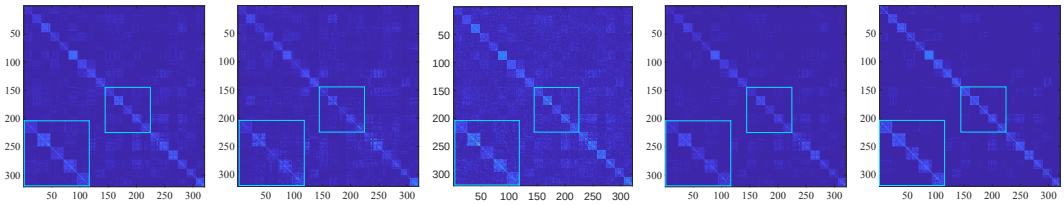
Note: "OT" indicates that the method exceeds the allowed running time.

$\{0.01, 0.03, 0.05, 0.07, 0.1, 1, 3\}$, $\tau \in \{0.01, 0.05, 0.07, 0.5, 1, 1.3\}$ and $\alpha \in [1 : 0.5 : 10]$, respectively.

4.2 Experimental Results

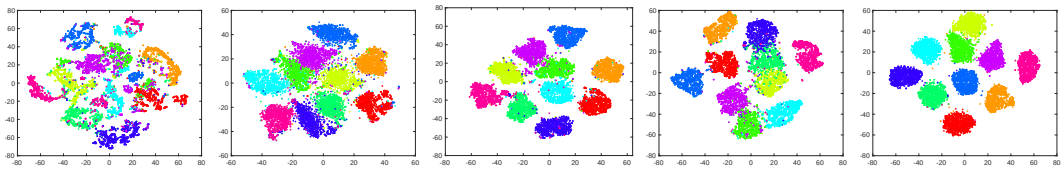
Tables 1-2 summarize the clustering performance of the proposed method against the compared methods on all benchmark datasets. The best scores are marked in bold, while the second best are underlined for ease of comparison. From these quantitative results, several observations can be made.

- Compared with convex low-rank regularization, nonconvex formulations generally achieve more favorable clustering performance. For example, on the BDGP dataset, LSGMC improves the ACC by 11.80% over the convex regularized-based MLRR method. This improvement can be ascribed to the greater flexibility of nonconvex regularizers in modeling singular values, which helps retain informative components and characterize the intrinsic low-rank structure more effectively.
- The proposed method achieves superior performance to the compared methods on all datasets. In particular, on the Hdigit dataset, our method surpasses the second-best competitor by 3.68%, 6.25%, and 7.53% in terms of ACC, NMI and ARI, respectively. These improvements mainly benefit from the unfolded network that enables joint optimization of the low-rank factors and structural constraints through back propagation.
- To evaluate the computational efficiency of DLRMF, we report the running times of all compared methods, as shown in Table 2. In most cases, DLRMF achieves the second-best efficiency, only slower than LargeMvC-Net, while remaining substantially faster than most conventional methods. These results provide compelling evidence of the computational efficiency of the proposed method.



(a) LSGMC (b) T-UMC (c) AMvSC (d) MLRR (e) Ours

Figure 1: Comparison of the affinity matrices constructed by different methods on the 100Leaves dataset (the first 20 subjects are shown).



(a) SMVAGC (b) TAD-MVC (c) CAL (d) LSGMC (e) Ours

Figure 2: The visualization of different methods on Hdigit dataset by t-SNE.

4.3 Qualitative Study

Affinity Matrix. Figure 1 provides a visual comparison of the affinity matrices learned by four competing methods and our method. As can be observed, LSGMC and T-UMC exhibit relatively diffuse off-diagonal responses, indicating substantial inter-cluster interference. AMvSC and MLRR also suffer from blurred block boundaries and pronounced spurious connections in the highlighted regions. In comparison, our method yields sharper and more compact diagonal blocks with a cleaner background. These observations suggest that our model preserves intra-cluster relationships more effectively while filtering out spurious correlations between different clusters.

Visualization of Embedded Representation. Figure 2 visualizes the clustering representations obtained by different methods on the Hdigit dataset using t-SNE, with colors indicating different categories. The competing methods in Figure 2(a)-(d) exhibit varying degrees of cluster overlap and sample dispersion, and several categories are not well separated in the embedding space. In contrast, our method produces a considerably cleaner structure, where samples from the same category are tightly grouped and distinct categories are separated by clear margins. These observations suggest that the proposed method learns a more discriminative and cluster-friendly representation, providing further evidence of its superior clustering capability.

4.4 Ablation Study

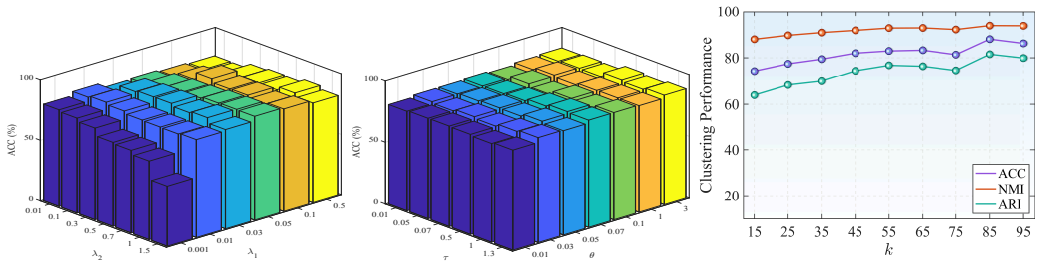
To further validate the clustering effectiveness of the proposed method, an ablation variant, termed DLRMF-1, is constructed by removing the trainable network and evaluated on four datasets. As shown in Table 3, DLRMF consistently outperforms DLRMF-1 while requiring noticeably less running time, with this efficiency advantage growing as the number of samples increases. The improvement stems from the proposed method updates parameters through backpropagation using the additionally introduced loss function, which guides the model toward more desirable outputs while avoiding SVD operations during optimization.

4.5 Parameter Sensitivity

To examine the sensitivity of λ_1, λ_2, τ, θ, and k, we vary these parameters via grid search and report the results on 100Leaves in Figure 3. It can be observed that the clustering performance remains consistently high across a broad range of parameter settings, with only limited variations. Moreover, the performance tends to improve as k increases and then gradually stabilizes. These

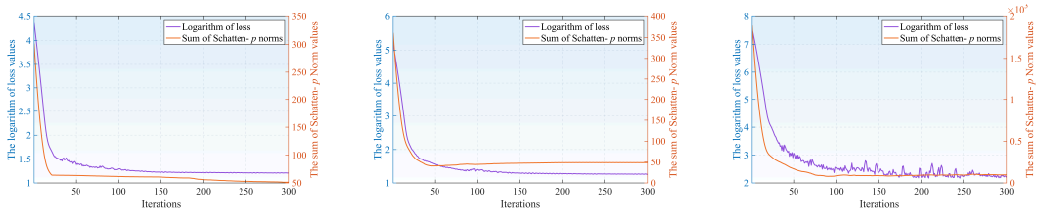
Table 3: Ablation comparison of DLRMF and DLRMF-1 on four datasets.

	BBCSport			BDGP			Wikipedia			Fashion		
Method	ACC	NMI	Time(s)	ACC	NMI	Time(s)	ACC	NMI	Time(s)	ACC	NMI	Time(s)
DLRMF-1	95.96	87.42	7.57	57.60	37.31	35.56	57.10	50.69	5.59	69.66	71.12	714.48
DLRMF	**97.06**	**90.17**	**5.81**	**64.32**	**40.48**	**10.07**	**61.06**	**52.55**	**1.33**	**80.37**	**77.66**	**62.44**



(a) (b) (c)

Figure 3: Parameter analysis of λ_1, λ_2, τ , θ and k in terms of clustering performance on 100leaves datasets. In each subfigure (a)(b)(c), the displayed parameters are varied while all other parameters are held fixed.



(a) 100Leaves (b) Wikipedia (c) Fashion

Figure 4: Convergence curves of the objective function and variation curves of the sum of the Schatten-p norms of all factors on three datasets.

results demonstrate that the proposed method is relatively insensitive to parameter changes and exhibits good robustness.

4.6 Convergence Analysis

Figure 4 reports the numerical convergence behavior of the proposed method on three datasets. The loss curves decrease rapidly at the beginning and gradually stabilize within a limited number of iterations, while the summed Schatten-p norms of all factors exhibit a similar decreasing trend, indicating that the learned representations are progressively encouraged toward low-rank structures. These results demonstrate that the proposed framework can effectively minimize the objective while promoting low-rank representations, thereby validating its effectiveness and convergence.

5 Conclusion

In this work, we proposed a novel MvSC approach named DLRMF. On the one hand, it formulates multi-factor, low-rank regularization and structural constraints within a unified framework to learn structured multiview representations. On the other hand, it transform the conventional iterative optimization procedure into differentiable learning steps, thereby avoiding repeated SVD operations. Then, an effective optimization algorithm with theoretical convergence analysis is developed for the resulting model. Extensive experiments on multiple benchmark datasets demonstrate the effectiveness and computational efficiency of the proposed framework.

REFERENCES

Wei Chang, Huimin Chen, Feiping Nie, Rong Wang, and Xuelong Li. Tensorized and compressed multi-view subspace clustering via structured constraint. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10434–10451, 2024.

Jie Chen, Shengxiang Yang, Hua Mao, and Conor Fahy. Multiview subspace clustering using low-rank representation. *IEEE Transactions on Cybernetics*, 52(11):12364–12378, 2022.

Siyu Chen, Er Wang, Lifan Peng, Yanchi Ou, Rong Tang, Xiaoqian Zhang, and Zhenwen Ren. Noise-invariant joint optimization for deep multi-view subspace clustering. *IEEE Transactions on Consumer Electronics*, 2026a.

Xu Chen, Zhiwen Yu, Ziwei Fan, Kaixiang Yang, and C. L. Philip Chen. Adaptive dictionary learning for multiview subspace clustering. *IEEE Transactions on Cybernetics*, 55(6):2833–2843, 2025a.

Xu Chen, Zhiwen Yu, Kaixiang Yang, and CL Philip Chen. A parameter-free multi-view clustering framework with adaptive anchors for large-scale data. *IEEE Transactions on Knowledge and Data Engineering*, 2026b.

Yawei Chen, Huibing Wang, Mingze Yao, Jinjia Peng, Guangqi Jiang, and Jiqing Zhang. Scalable multi-view clustering based on tight anchor distribution. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 7739–7747, 2025b.

Zhaoliang Chen, Jie Yao, Guobao Xiao, and Shiping Wang. Efficient and differentiable low-rank matrix completion with back propagation. *IEEE Transactions on Multimedia*, 25:228–242, 2023.

Zhe Chen, Fanhui Meng, Tianyang Xu, and Xiao-Jun Wu. Cluster-aware anchor learning for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17714–17723, 2026c.

Meng Ding, Jing-Hua Yang, Xi-Le Zhao, Jie Zhang, and Michael K. Ng. Nonconvex low-rank tensor representation for multi-view subspace clustering with insufficient observed samples. *IEEE Transactions on Knowledge and Data Engineering*, 37(6):3583–3597, 2025.

Shide Du, Chunming Wu, Zihan Fang, Wendi Zhao, Yilin Wu, Changwei Wang, and Shiping Wang. Largemvc-net: Anchor-based deep unfolding network for large-scale multi-view clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 1714–1723, 2025.

Yao Fu, Dong Hu, and Zhi Wang. Nonconvex low-rank tensor representation with deep priors for multiview subspace clustering. In *Forty-third International Conference on Machine Learning*, 2026a.

Yao Fu, Zhi Wang, Dong Hu, Tao Jia, Chao Gao, and Zhen Wang. A unified framework with capped tensor norm minimization for multiview subspace learning. *IEEE Transactions on Knowledge and Data Engineering*, 38(9):5607–5621, 2026b.

Wenbin Gao, Hua Meng, Zhengchun Zhou, Ming Li, and Ruibin Ren. Adaptive multi-view subspace clustering with dynamic prior update. *Applied Soft Computing*, pp. 114848, 2026.

Zhibin Gu and Songhe Feng. From dictionary to tensor: A scalable multi-view subspace clustering framework with triple information enhancement. In *Advances in Neural Information Processing Systems*, volume 37, pp. 103545–103573. Curran Associates, Inc., 2024.

Jipeng Guo, Yanfeng Sun, Junbin Gao, Yongli Hu, and Baocai Yin. Logarithmic schatten-p norm minimization for tensorial multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3396–3410, 2023.

Dongdong Hou, Yang Cong, Gan Sun, Jiahua Dong, Jun Li, and Kai Li. Fast multi-view outlier detection via deep encoder. *IEEE Transactions on Big Data*, 8(4):1047–1058, 2020.

Wei Lan, Tianchuan Yang, Qingfeng Chen, Shichao Zhang, Yi Dong, Huiyu Zhou, and Yi Pan. Multiview subspace clustering via low-rank symmetric affinity graph. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8):11382–11395, 2024.

Jia-Qi Lin, Man-Sheng Chen, Chang-Dong Wang, and Haizhang Zhang. A tensor approach for uncoupled multiview clustering. *IEEE Transactions on Cybernetics*, 54(2):1236–1249, 2024.

Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

Suyuan Liu, Qing Liao, Siwei Wang, Xinwang Liu, and En Zhu. Robust and consistent anchor graph learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36 (8):4207–4219, 2024.

Zhuo Liu, Dong Hu, and Zhi Wang. Efficient implementation of truncated frobenius norm minimization for latent low-rank representation. *IEEE Transactions on Signal and Information Processing over Networks*, 11:1394–1408, 2025.

Zhuo Liu, Dong Hu, and Zhi Wang. Nonconvex low-rank tensor representation with self-enhanced structure penalty for multi-view subspace clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2026.

Chao Luo, Jie Zhang, and Xiaoqian Zhang. Tensor multi-view clustering method for natural image segmentation. *Expert Systems with Applications*, 260:125431, 2025.

Huimin Ma, Siwei Wang, Junpu Zhang, Shengju Yu, Suyuan Liu, Xinwang Liu, and Kunlun He. Symmetric multi-view subspace clustering with automatic neighbor discovery. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(9):8766–8778, 2024.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

Feiping Nie, Heng Huang, and Chris H. Q. Ding. Low-rank matrix recovery via efficient schatten p-norm minimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 655–661, 2012.

Yalan Qin and Guorui Feng. Unified and efficient multi-view clustering from probabilistic perspective. In *International Conference on Learning Representations*, volume 2026, pp. 101262–101274, 2026.

Yalan Qin, Nan Pu, Guorui Feng, and Nicu Sebe. Robust consensus anchor learning for efficient multi-view subspace clustering. In *Forty-second International Conference on Machine Learning*, 2025.

Yiwen Shan, Dong Hu, Zhi Wang, and Tao Jia. Multi-channel nuclear norm minus frobenius norm minimization for color image denoising. *Signal Processing*, 207:108959, 2023.

Fanhua Shang, Yuanyuan Liu, and James Cheng. Scalable algorithms for tractable schatten quasi-norm minimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 2016–2022, 2016.

Fanhua Shang, James Cheng, Yuanyuan Liu, Zhi-Quan Luo, and Zhouchen Lin. Bilinear factor matrix norm minimization for robust pca: Algorithms and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2066–2080, 2018.

Gábor Takács, István Pilászy, Bottyán Németh, and Domonkos Tikk. Scalable collaborative filtering approaches for large recommender systems. *Journal of Machine Learning Research*, 10:623–656, 2009.

Yuze Tan, Yixi Liu, Shudong Huang, Wentao Feng, and Jiancheng Lv. Sample-level multi-view graph clustering. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23966–23975, 2023.

Yongqiang Tang, Yuan Xie, and Wensheng Zhang. Affine subspace robust low-rank self-representation: From matrix to tensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):9357–9373, 2023.

Kim-Chuan Toh and Sangwoon Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of optimization*, 6(615-640), 2010.

Xinhang Wan, Xinwang Liu, Jiyuan Liu, Siwei Wang, Yi Wen, Weixuan Liang, En Zhu, Zhe Liu, and Lu Zhou. Auto-weighted multi-view clustering for large-scale data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

Qianqian Wang, Zihao Zhang, Wei Feng, Zhiqiang Tao, and Quanxue Gao. Contrastive multi-view subspace clustering via tensor transformers autoencoder. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 21207–21215, 2025a.

Shuqin Wang, Zhiping Lin, Qi Cao, Yigang Cen, and Yongyong Chen. Bi-nuclear tensor schatten-p norm minimization for multi-view subspace clustering. *IEEE Transactions on Image Processing*, 32:4059–4072, 2023.

Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE Transactions on Image Processing*, 33: 4627–4639, 2024a.

Yuefei Wang, Jiaying Mei, Zixu Wang, Jinyan Cao, Xinyuan Wang, Siyi Qiu, Binxiong Li, and Xi Yu. Multi-view outlier detection via tensor decomposition and information decoupling. *Applied Intelligence*, 56(10):341, 2026.

Zhi Wang, Chao Gao, Xiaohu Luo, Ming Tang, Jianjun Wang, and Wu Chen. Accelerated inexact matrix completion algorithm via closed-form q-thresholding ($q = 1/2, 2/3$) operator. *Int. J. Mach. Learn. Cybern.*, 11(10):2327–2339, 2020.

Zhi Wang, Zhuo Liu, Dong Hu, and Tao Jia. Nonconvex multiview subspace clustering framework with efficient method designs and theoretical analysis. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 5162–5170, 2024b.

Zhi Wang, Dong Hu, Zhuo Liu, Chao Gao, and Zhen Wang. Iteratively capped reweighting norm minimization with global convergence guarantee for low-rank matrix learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(3):1923–1940, 2025b.

Xiaolin Xiao, Yue-Jiao Gong, and Yicong Zhou. Learning orthogonal latent representations for multi-view clustering. *IEEE Transactions on Multimedia*, 27:8565–8578, 2025.

Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision*, 126(11):1157–1179, 2018.

Chen Xu, Zhouchen Lin, and Hongbin Zha. A unified convex surrogate for the schatten-p norm. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 926–932, 2017.

Tinghe Yan, Qiang Guo, Jian-Xun Mi, and Weisheng Li. Multi-view subspace clustering via tensor nuclear norm factorization. *Pattern Recognition*, 174:112969, 2026.

Junfeng Yang, Wotao Yin, Yin Zhang, and Yilun Wang. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM Journal on Imaging Sciences*, 2(2):569–592, 2009.

Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Deep admm-net for compressive sensing mri. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 10–18, 2016.

Yongbo Yu, Zhoumin Lu, Feiping Nie, Weizhong Yu, Zongcheng Miao, and Xuelong Li. Pseudo-label guided bidirectional discriminative deep multi-view subspace clustering. *IEEE Transactions on Knowledge and Data Engineering*, 37(7):4213–4224, 2025.

Hengmin Zhang, Jian Yang, Fanhua Shang, Chen Gong, and Zhenyu Zhang. LRR for subspace segmentation via tractable schatten- p norm minimization and factorization. *IEEE Transactions on Cybernetics*, 49(5):1722–1734, 2019.

Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837, 2018.

Zihao Zhao, Kai-Chia Mo, Shing-Hei Ho, Brandon Amos, and Kai Wang. A fully first-order layer for differentiable optimization. In *Forty-third International Conference on Machine Learning*, 2026.

Haochen Zhou, Guofeng Ding, Mouxing Yang, Peng Hu, Yijie Lin, and Xi Peng. Uncover underlying correspondence for robust multi-view clustering. In *The Fourteenth International Conference on Learning Representations*, 2026.

A Appendix

This section presents the full optimization derivations and convergence proof, together with additional experimental results.

A.1 LLM USAGE

The use of LLMs is limited to language editing for greater clarity and fluency, without being employed for any additional purpose.

A.2 Nonconvex Schatten-p Regularization

Given the limited approximation capability of the nuclear norm, the Schatten- p norm has attracted considerable attention as a more flexible surrogate for matrix rank. The Schatten-p norm of matrix Z can be defined as

$$\|Z\|_{S_p} = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i(Z)^p \right)^{\frac{1}{p}} = \left(\mathrm{Tr}\left((Z^T Z)^{\frac{p}{2}} \right) \right)^{\frac{1}{p}}, \tag{20}$$

where $0 < p < \infty$, and $\sigma_i(Z)$ denotes the i-th singular value associated with $Z \in \mathbb{R}^{n\times n}$. Generally, the minimization problem for the Schatten-p norm can be solved via the following generalized thresholding operator Zhang et al. (2019)

$$\mathrm{Prox}_{S_p}(Y) = \arg\min_{Z} \frac{1}{2}\|Z - Y\|_F^2 + \tau \|Z\|_{S_p}^p. \tag{21}$$

However, evaluating this operator generally requires SVD on the coefficient matrix, leading to considerable computational cost for large-scale data. To alleviate computational burden, several studies have explored matrix-factorization-based surrogates for specific choices of p Takács et al. (2009); Shang et al. (2016; 2018).

$$\begin{aligned} \|Z\|_* &= \min_{Z=UV^T} \frac{1}{2}\left(\|U\|_F^2 + \|V\|_F^2 \right), \\ 2\|Z\|_{S_{1/2}}^{1/2} &= \min_{Z=UV^T} \left(\|U\|_* + \|V\|_* \right), \\ \frac{3}{2}\|Z\|_{S_{2/3}}^{2/3} &= \min_{Z=UV^T} \left(\|U\|_* + \frac{1}{2}\|V\|_F^2 \right). \end{aligned} \tag{22}$$

Building on these prior works, the multi-Schatten-p norm surrogate (MSS) was studied to unify various decomposition-based surrogates Xu et al. (2017). Consider a matrix Z factorized into I factors as $Z = \prod_{i=1}^{I} Z_i$, where $Z_1 \in \mathbb{R}^{m\times d_1}$, $Z_i \in \mathbb{R}^{d_{i-1}\times d_i}$ for $i = 2, \ldots, I-1$, and $Z_I \in \mathbb{R}^{d_{I-1}\times n}$. For $\mathrm{rank}(Z) = r \le \min\{d_i,\ i = 1, \ldots, I\}$, the MSS function can be expressed as

$$\frac{1}{p}\|Z\|_{S_p}^p = \min_{Z=\prod_{i=1}^{I} Z_i} \sum_{i=1}^{I} \frac{1}{p_i}\|Z_i\|_{S_{p_i}}^{p_i}, \quad \frac{1}{p} = \sum_{i=1}^{I} \frac{1}{p_i}. \tag{23}$$

Accordingly, the original low-rank optimization can be decomposed into a set of factor-wise subproblems, providing a more flexible formulation for subsequent optimization. The MSS objective is optimized in a block-wise manner via BCD, where one factor matrix is updated while the others are kept fixed. For each block, PALM is employed to obtain the proximal-gradient update, yielding the following subproblem:

$$\begin{aligned} Z_i^{(k+1)} &= \arg\min_{Z_i} \; f\left(Z_i^{(k)}\right) + \left\langle \nabla f\left(Z_i^{(k)}\right), Z_i - Z_i^{(k)} \right\rangle + \frac{L_i^{(k-1)}}{2} \left\| Z_i - Z_i^{(k)} \right\|_F^2 + \frac{1}{p_i} \left\| Z_i \right\|_{S_{p_i}}^{p_i} \\ &= \arg\min_{Z_i} \; \frac{L_i^{(k-1)}}{2} \left\| Z_i - Q_i \right\|_F^2 + \frac{1}{p_i} \left\| Z_i \right\|_{S_{p_i}}^{p_i}, \end{aligned} \tag{24}$$

where $Q_i = Z_i^{(k)} - \frac{1}{L_i^{(k-1)}} \nabla f(Z_i^{(k)})$, and $L_i^{(k-1)}$ denotes the Lipschitz constant of $\nabla f(\cdot)$, satisfying

$$\left\| \nabla f(Z_i) - \nabla f(\widetilde{Z}_i) \right\|_F \leq L_i^{(k-1)} \left\| Z_i - \widetilde{Z}_i \right\|_F. \tag{25}$$

The subproblem in Eq. 24 can be solved through the proximal operator associated with the Schatten-p regularizer, given by

$$Z_i^{(k+1)} = \mathrm{Prox}_{S_{p_i}} \left(Z_i^{(k)} - \frac{1}{L_i^{(k-1)}} \nabla f\left(Z_i^{(k)}\right) \right). \tag{26}$$

Nevertheless, SVD computations still involved in solving the MSS-based optimization problem incur substantial computational overhead.

A.3 Proof of Theorem 1

Proof. Because $Z_i^{(v)} = \frac{1}{2} W_i^{(v,k)} (P_i^{(v)})^\dagger$ satisfies the first-order optimality condition of the weighted low-rank formulation, it is the solution to the following weighted quadratic subproblem:

$$\arg\min_{Q_i^{(v)}} \mathrm{Tr}\left((Q_i^{(v)})^T Q_i^{(v)} P_i^{(v)} \right), \quad s.t. Q_i^{(v)} = Z_i^{(v)} \tag{27}$$

where $P_i^{(v)} = \frac{1}{2} \left((Q_i^{(v)})^T Q_i^{(v)} \right)^{\frac{p_i-2}{2}}$. Accordingly, it is satisfied that

$$\mathrm{Tr}\left((Z_i^{(v)})^T Z_i^{(v)} P_i^{(v)} \right) \leq \mathrm{Tr}\left((Q_i^{(v)})^T Q_i^{(v)} P_i^{(v)} \right). \tag{28}$$

By substituting $P_i^{(v)} = \frac{1}{2} \left((Q_i^{(v)})^T Q_i^{(v)} \right)^{\frac{p_i-2}{2}}$ into Eq. (28), we get

$$\frac{p_i}{2} \mathrm{Tr}\left((Z_i^{(v)})^T Z_i^{(v)} \left((Q_i^{(v)})^T Q_i^{(v)} \right)^{\frac{p_i-2}{2}} \right) \leq \frac{p_i}{2} \mathrm{Tr}\left((Q_i^{(v)})^T Q_i^{(v)} \left((Q_i^{(v)})^T Q_i^{(v)} \right)^{\frac{p_i-2}{2}} \right). \tag{29}$$

Let $A = (Z_i^{(v)})^T Z_i^{(v)}$ and $A_t = (Q_i^{(v)})^T Q_i^{(v)}$. According to Lemma 1, we have

$$\mathrm{Tr}\left(A^{\frac{p_i}{2}} \right) - \frac{p_i}{2} \mathrm{Tr}\left(A A_t^{\frac{p_i-2}{2}} \right) \leq \mathrm{Tr}\left(A_t^{\frac{p_i}{2}} \right) - \frac{p_i}{2} \mathrm{Tr}\left(A_t A_t^{\frac{p_i-2}{2}} \right). \tag{30}$$

Combining Eq. (29) and Eq. (30), we obtain

$$\mathrm{Tr}\left(\left((Z_i^{(v)})^T Z_i^{(v)} \right)^{\frac{p_i}{2}} \right) \leq \mathrm{Tr}\left(\left((Q_i^{(v)})^T Q_i^{(v)} \right)^{\frac{p_i}{2}} \right). \tag{31}$$

According to the definition of the Schatten-p norm Eq. (20), it follows that

$$\left\| Z_i^{(v)} \right\|_{S_{p_i}}^{p_i} \leq \left\| Q_i^{(v)} \right\|_{S_{p_i}}^{p_i}, \tag{32}$$

which completes the proof when $0 < p_i \leq 2$. □

Algorithm 1 Multi-Factor Low-Rank Unfolding Network (DLRMF)

Input: Multi-view data $\{X^{(v)}\}_{v=1}^{V} \in \mathbb{R}^{d_v \times n}$, Schatten-$p$ indicators $\{p_i\}_{i=1}^{I}$.
Output: The self-representation matrix $\{Z^{(v)}\}_{v=1}^{V}$ for each view.

Initialize learnable weight matrices $\{W_i^{(v)}\}_{i=1}^{I}$ and step parameters $\{\rho_i^{(v)}\}_{i=1}^{I}$,$v = 1, ..., V$;
Initialize low-dimensional factors $\{Z_i^{(v)}\}_{i=1}^{I}$, sparse matrices $E^{(v)}$,$v = 1, ..., V$ and counter $t = 0$;
while $\mathcal{J}$ has not converged and the early-stop condition is not met **do**
    for $v = 1 \to V$ **do**
        for $i = 1 \to I$ **do**
            Obtain the Lipschitz constant $L_i^{(v)}$ by Eq. (25);
            Calculate the low rank operator input $Q_i^{(v)}$ by Eq. (14);
            Update $Z_i^{(v)}$ using the differentiable low-rank operator Eq. (13);
        end for
        Update $E^{(v)}$ by Eq. (17);
    end for
    Update counter $t = t + 1$;
    Update weight matrices $W_i^{(v)}$ and step parameters ρ_i through backpropagation;
end while
return $\{Z^{(v)}\}_{v=1}^{V}$.

Algorithm 2 Clustering with Learned Self-Representations matrices

Input: Learned self-representation matrices $\{Z^{(v)}\}_{v=1}^{V}$, number of clusters k, and parameter α.
Output: Clustering partition $\{C_j\}_{j=1}^{k}$.

Aggregate the view-specific representations as $Z^* = \sum_{v=1}^{V} Z^{(v)}$.
Compute the compact SVD of Z^* as $Z^* = U^* \Sigma^* (V^*)^T$.
Construct the spectral embedding $M = U^* (\Sigma^*)^{1/2}$.
Generate the affinity matrix W_n with

$$[W_n]_{ij} = \left[\max \left(\frac{m_i^T m_j}{\|m_i\|_2 \|m_j\|_2}, 0 \right) \right]^{2\alpha}, \tag{33}$$

where m_i denotes the i-th row of M.
Form the normalized Laplacian $L = D^{-1/2}(D - W_n)D^{-1/2}$, where D is the degree matrix of W.
Extract the k eigenvectors associated with the smallest k eigenvalues of L and stack them into $V \in \mathbb{R}^{n \times k}$.
Regard the i-th row of V as the spectral representation $y_i \in \mathbb{R}^k$ of the i-th sample.
Apply k-means to $\{y_i\}_{i=1}^{n}$ and obtain the partition $\{C_j\}_{j=1}^{k}$.
return $\{C_j\}_{j=1}^{k}$.

The above result proves the convergence property of the Schatten-p_i regularization for each factor-update block. Since every factor in each view is updated toward a lower Schatten-p_i value, the overall low-rank regularization formed by these factor-wise terms is progressively reduced. Moreover, since the overall low-rank regularization is non-negative and thus bounded from below, which guarantees its convergence during the optimization of DLRMF.

A.4 Back Propagation for DLRMF

In this subsection, we derive the gradients required for optimizing the trainable variables in the factor-update blocks. Since all view-specific factor blocks share the same computational procedure, it is sufficient to consider the i-th factor associated with the v-th view. The gradients of the learnable weight matrix $W_i^{(v,k)}$ and the step-control parameter $\rho_i^{(v)}$ are derived as follows.

1) Gradient with respect to $W_i^{(v,k)}$. According to Eq. (13), the output of the i-th differentiable low-rank operator is

$$Z_i^{(v)} = \frac{1}{2} W_i^{(v,k)} \left(P_i^{(v)}\right)^{\dagger}. \tag{34}$$

Therefore, applying the matrix chain rule gives

$$\frac{\partial \mathcal{J}}{\partial W_i^{(v,k)}} = \frac{1}{2} \frac{\partial \mathcal{J}}{\partial Z_i^{(v)}} \frac{\partial Z_i^{(v)}}{\partial W_i^{(v,k)}} = \frac{1}{2} \frac{\partial \mathcal{J}}{\partial Z_i^{(v)}} \left[\left(P_i^{(v)}\right)^{\dagger}\right]^T. \tag{35}$$

Here, $\frac{\partial \mathcal{J}}{\partial Z_i^{(v)}}$ is obtained from Eq. (15), where the gradients are propagated from the reconstruction, cross-view consistency, and symmetry terms.

2) Gradient with respect to $\rho_i^{(v)}$. The step-control parameter $\rho_i^{(v)}$ affects the objective through the intermediate variable $Q_i^{(v)}$. Hence, its gradient can be written as

$$\frac{\partial \mathcal{J}}{\partial \rho_i^{(v)}} = \mathrm{Tr}\left[\left(\frac{\partial \mathcal{J}}{\partial Q_i^{(v)}}\right)^T \frac{\partial Q_i^{(v)}}{\partial \rho_i^{(v)}}\right]. \tag{36}$$

From 14,

$$Q_i^{(v)} = Z_i^{(v)} - \frac{1}{\rho_i^{(v)} L_i^{(v)}} \nabla_{Z_i^{(v)}} F, \tag{37}$$

and consequently,

$$\frac{\partial Q_i^{(v)}}{\partial \rho_i^{(v)}} = \frac{1}{\left(\rho_i^{(v)}\right)^2 L_i^{(v)}} \nabla_{Z_i^{(v)}} F. \tag{38}$$

Substituting Eq. (38) into Eq. (36) yields

$$\frac{\partial \mathcal{J}}{\partial \rho_i^{(v)}} = \frac{1}{\left(\rho_i^{(v)}\right)^2 L_i^{(v)}} \mathrm{Tr}\left[\left(\frac{\partial \mathcal{J}}{\partial Q_i^{(v)}}\right)^T \nabla_{Z_i^{(v)}} F\right]. \tag{39}$$

To obtain $\frac{\partial \mathcal{J}}{\partial Q_i^{(v)}}$, we further propagate the gradient through the factor-wise low-rank operator. Let

$$S_i^{(v)} = \left(P_i^{(v)}\right)^{\dagger}. \tag{40}$$

From $Z_i^{(v)} = \frac{1}{2} W_i^{(v,k)} S_i^{(v)}$, we have

$$\frac{\partial \mathcal{J}}{\partial S_i^{(v)}} = \frac{1}{2} \left(W_i^{(v,k)}\right)^T \frac{\partial \mathcal{J}}{\partial Z_i^{(v)}}. \tag{41}$$

Following the differentiable inverse mapping used in the factor-update operator, the gradient with respect to $P_i^{(v)}$ is expressed as

$$\frac{\partial \mathcal{J}}{\partial P_i^{(v)}} = -\left(S_i^{(v)}\right)^T \frac{\partial \mathcal{J}}{\partial S_i^{(v)}} \left(S_i^{(v)}\right)^T. \tag{42}$$

Combining Eq. (41) with Eq. (42) gives

$$\frac{\partial \mathcal{J}}{\partial P_i^{(v)}} = -\frac{1}{2} \left(S_i^{(v)}\right)^T \left(W_i^{(v,k)}\right)^T \frac{\partial \mathcal{J}}{\partial Z_i^{(v)}} \left(S_i^{(v)}\right)^T. \tag{43}$$

Recalling that

$$P_i^{(v)} = \frac{1}{2} \left(\left(Q_i^{(v)}\right)^T Q_i^{(v)}\right)^{\frac{p_i - 2}{2}}, \tag{44}$$

Table 5: Clustering performance comparison in terms of F-score, Precision, and Recall.

Dataset	Metrics	CAL	SMVAGC	TAD-MVC	TC-MVSC	LargeMvC-Net	MLRR	SMSC-AND	NLRSC	AMvSC	SLMVGC	LSGMC	T-UMC	OLR-MVC	Ours
100Leaves	F-score	74.54	50.54	53.54	6.07	63.45	81.68	50.16	75.16	77.73	56.50	68.60	45.71	75.60	**82.32**
	Precision	69.31	39.53	50.39	3.57	58.71	75.78	47.12	75.40	77.34	54.33	75.00	43.31	71.21	**79.52**
	Recall	82.30	70.44	53.07	51.12	69.04	82.47	53.63	73.14	**85.33**	58.86	83.81	42.75	79.46	**85.33**
BDGP	F-score	43.85	34.85	40.64	31.28	40.12	46.73	48.57	37.94	50.55	33.55	46.03	23.89	45.09	**50.64**
	Precision	41.95	32.78	36.24	22.26	38.03	48.77	44.65	37.50	**51.36**	29.48	34.91	23.71	44.41	48.96
	Recall	45.94	37.41	46.24	38.76	42.57	45.64	52.04	38.38	**54.00**	38.93	51.65	4.50	45.78	52.43
BBCSport	F-score	85.83	83.04	66.13	70.78	83.14	90.85	51.56	87.13	93.12	59.32	88.00	76.51	93.25	**94.15**
	Precision	87.96	83.44	64.72	60.84	78.57	91.21	45.60	88.42	93.72	56.60	91.50	69.46	**94.28**	93.88
	Recall	90.20	82.86	67.61	84.60	88.28	91.27	60.49	92.18	90.91	62.36	94.17	58.71	92.24	**94.42**
Wikipedia	F-score	21.58	48.24	19.13	39.62	42.56	49.84	27.87	50.55	13.38	42.68	13.31	47.18	45.57	**52.46**
	Precision	19.96	48.63	19.57	34.02	43.48	48.74	28.74	**50.83**	21.81	42.60	12.06	48.09	32.86	49.44
	Recall	23.93	47.88	18.70	47.42	41.68	48.97	27.05	50.27	16.15	42.75	14.88	41.96	**74.31**	55.88
MNIST	F-score	82.00	79.41	76.19	81.29	64.33	61.40	61.08	69.52	70.18	84.07	69.91	77.93	76.25	**86.02**
	Precision	81.40	78.84	74.87	**89.90**	62.71	61.12	59.24	61.83	62.41	83.62	62.26	68.82	82.80	85.63
	Recall	82.62	79.99	77.55	82.71	66.04	67.39	63.04	79.39	80.17	84.53	79.71	56.75	70.66	**86.42**
Hdigit	F-score	87.82	55.99	86.03	59.02	57.04	36.08	OT	76.55	90.95	68.70	77.62	OT	50.13	**97.72**
	Precision	87.75	53.66	85.87	48.85	55.17	36.78	OT	73.60	90.87	67.01	72.66	OT	37.10	**97.71**
	Recall	87.90	58.59	86.19	74.55	59.06	29.30	OT	79.74	91.03	70.48	83.29	OT	77.29	**97.73**
Fashion	F-score	62.32	65.65	64.00	40.09	61.81	OT	OT	62.09	58.25	48.77	70.90	OT	51.21	**72.22**
	Precision	59.61	62.30	61.19	25.97	59.23	OT	OT	39.12	54.31	45.04	67.31	OT	35.59	**68.19**
	Recall	65.40	71.09	67.07	67.89	64.65	OT	OT	74.32	62.80	53.17	74.91	OT	43.16	**76.75**
NUSWIDEOBJ	F-score	8.76	8.48	8.04	OT	9.10	OT	OT	OT	OT	10.20	8.27	OT	OT	**12.81**
	Precision	8.29	8.53	7.96	OT	**12.36**	OT	OT	OT	OT	7.89	7.86	OT	OT	9.37
	Recall	10.22	6.71	6.36	OT	7.20	OT	OT	OT	OT	8.13	8.77	OT	OT	**20.28**

Note: "OT" indicates that the method exceeds the allowed running time.

the gradient propagated to $Q_i^{(v)}$ can be written in matrix form as

$$\frac{\partial \mathcal{J}}{\partial Q_i^{(v)}} = \frac{p_i - 2}{2} Q_i^{(v)} \frac{\partial \mathcal{J}}{\partial P_i^{(v)}} \left(\left(Q_i^{(v)} \right)^T Q_i^{(v)} \right)^{\frac{p_i - 4}{2}}. \tag{45}$$

Therefore, Eq. (35)–Eq. (45) establish the complete gradient propagation paths for $W_i^{(v,k)}$ and $\rho_i^{(v)}$. The resulting gradients are passed to the optimizer to jointly learn the factor-wise weight matrices and adaptive step-control parameters during network training.

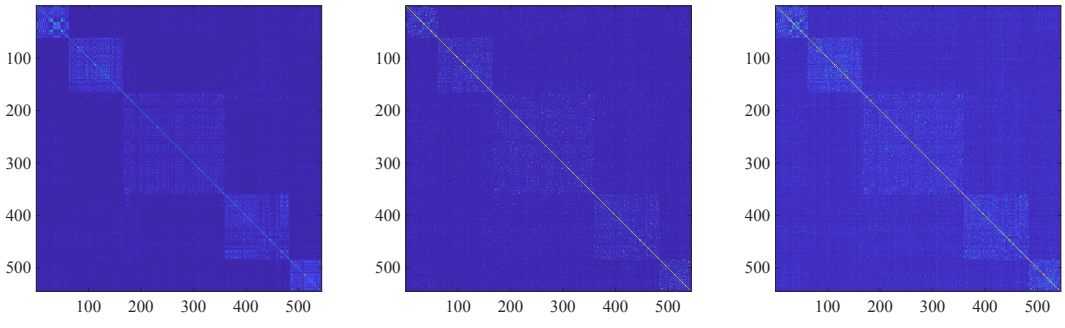
A.5 SUPPLEMENTARY EXPERIMENTS

This section provides additional experimental results that the main text does not report. Table 4 presents detailed information on the datasets used in the experiments. Table 5 records the F-score, Precision, and Recall metrics of thirteen comparison methods and our proposed method on all datasets. It can be seen that our method achieves best or second best in most cases, achieving satisfactory performance. Figure 5 and Figure 6 present the affinity graphs and t-SNE visualization generated by different methods on the BBCSport dataset. Figure 7 shows the t-SNE visualizations of different methods on the MNIST4 dataset. These results collectively demonstrate that the proposed method achieves highly competitive clustering performance. Figure 8 illustrates the convergence curves of the loss function and the Schatten-p norm values of the proposed method on the remaining datasets, further confirming the strong convergence of our method. Figure 11 presents the running time comparison between our method and all competing methods on the remaining four datasets.

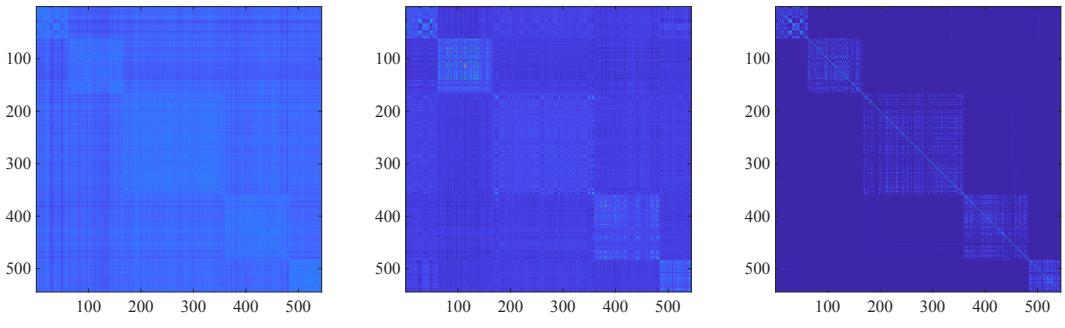
Table 6 summarizes the search ranges of the main parameters for all compared methods. For each parameter, only the minimum and maximum values of the corresponding search range are reported for clarity and compactness, where c denotes the number of clusters in each dataset.

Table 4: Statistical features for eight datasets.

Dataset	Sample	View/Cluster	Dimension
BBCSport	544	2 / 5	3183/3203
100leaves	1600	3 / 100	64/64/64
BDGP	2500	3 / 5	250/500/1000
Wikipedia	2866	2 / 10	128/10
MNIST4	4000	3 / 4	30/9/30
Fashion	10000	3 / 10	784/784/784
Hdigit	10000	2 / 10	784/256
NUSWIDEOBJ	30000	5 / 31	65/226/145/74/129

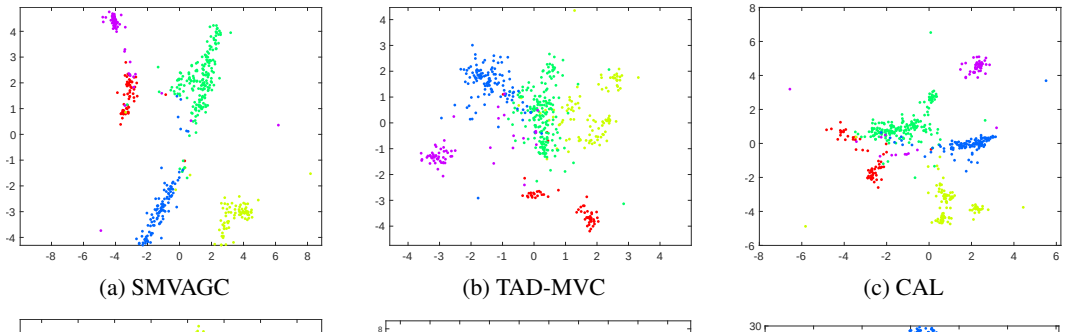


(a) LSGMC (b) T-UMC (c) AMvSC

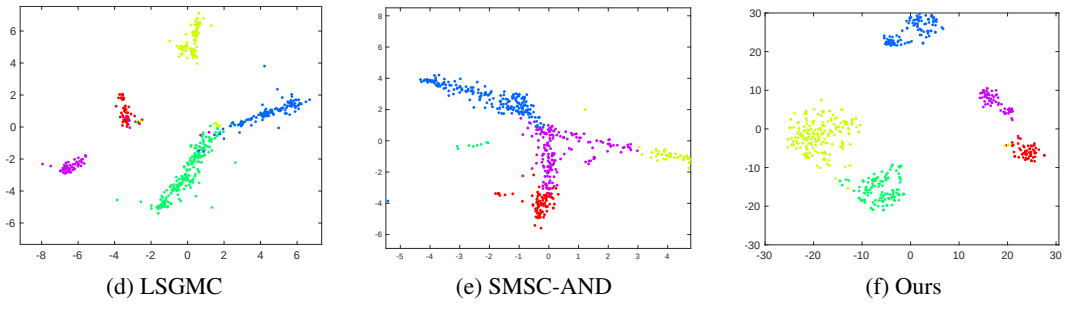


(d) MLRR (e) CAL (f) Ours

Figure 5: Comparison of the affinity matrices constructed by different methods on the BBCSport dataset.

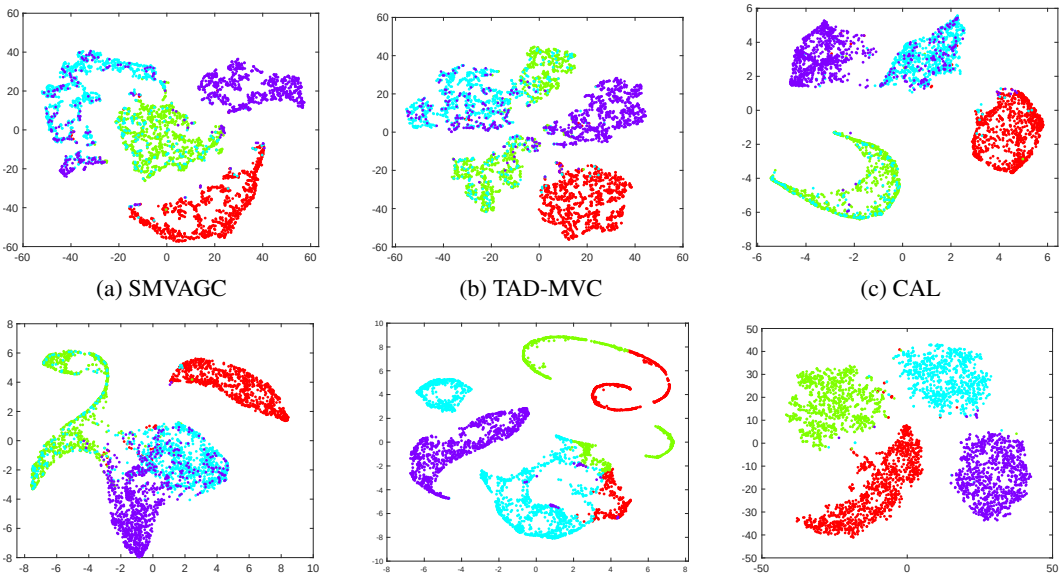


(a) SMVAGC (b) TAD-MVC (c) CAL



(d) LSGMC (e) SMSC-AND (f) Ours

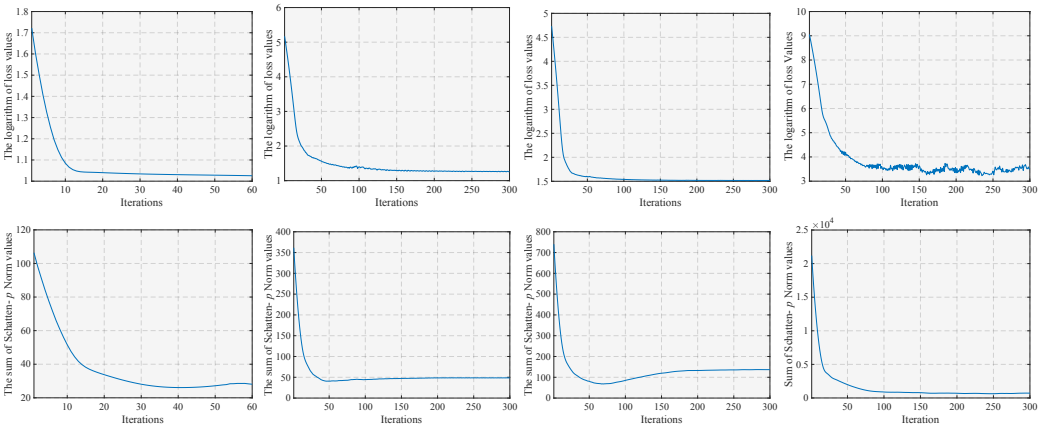
Figure 6: The t-SNE visualizations of different methods on the BBCSport dataset.



(a) SMVAGC (b) TAD-MVC (c) CAL

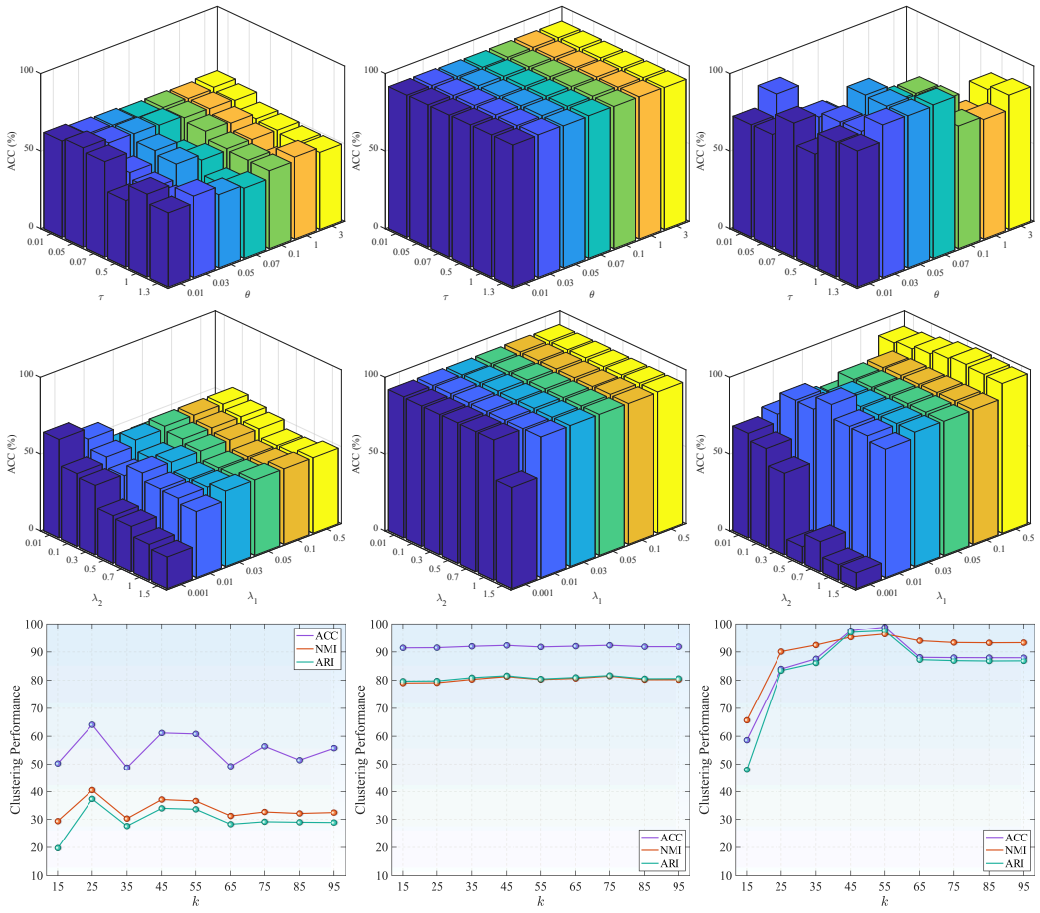
(d) LSGMC (e) SMSC-AND (f) Ours

Figure 7: The t-SNE visualizations of different methods on the MNIST4 dataset.



(a) BBCSport (b) Wikipedia (c) BDGP (d) NUSWIDEOBJ

Figure 8: Convergence curves of the objective function (the first row) and variation curves of the sum of the Schatten-p Norms of all factors (the second row) on four datasets.



(a) BDGP($\alpha = 2.5$) (b) MNIST4 ($\alpha = 10$) (c) Hdigit ($\alpha = 6$)

Figure 9: Parameter analysis of λ_1, λ_2, τ , θ and k in terms of ACC over BDGP, MNIST4 and Hdigit datasets.

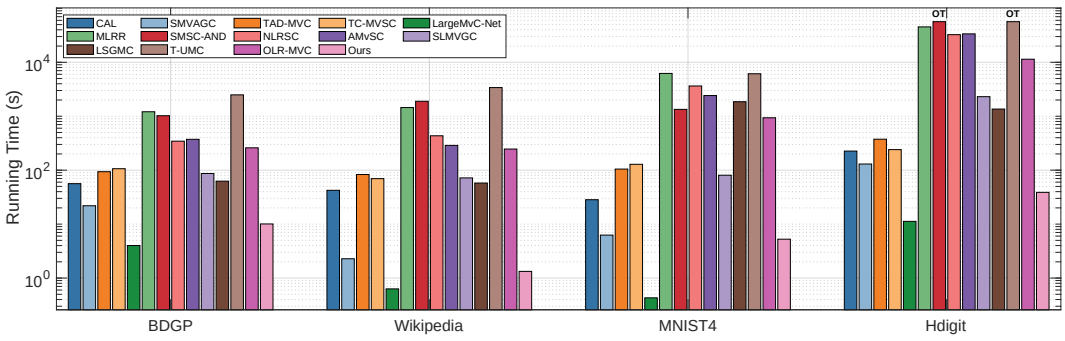


Figure 10: The running time of all methods and the proposed method on BDGP, Wikipedia, MNIST4 and Hdigit datasets.

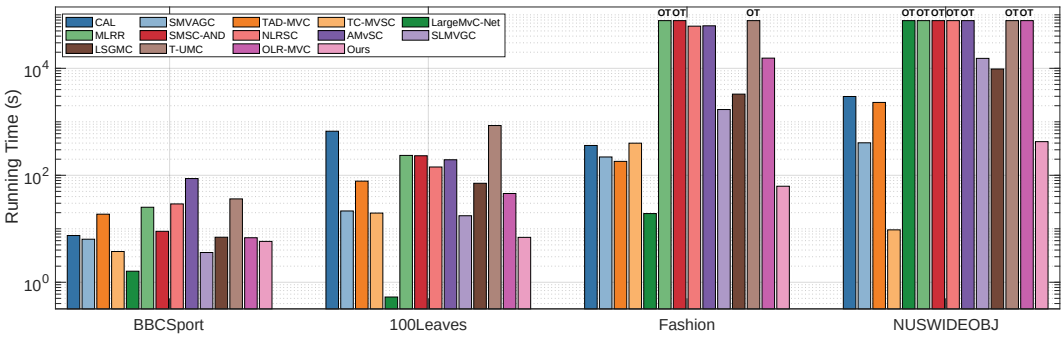


Figure 11: The running time of all methods and the proposed method on remaining BBCSport, 100leaves, Fashion and NUSWIDEOBJ datasets.

Table 6: Parameter search ranges of the compared methods.

Method	Parameter 1	Parameter 2	Parameter 3	Parameter 4
CAL	[1e-4,1e4]	[1e-4,1e4]	[1e-4,1e4]	–
RCAGL	[c,3c]	[0,1e6]	–	–
SMVAGC	[c,5c]	[1,1e3]	–	–
TAD-MVC	–	–	–	–
TC-MVSC	[1e-3,2]	[0.1,0.9]	–	–
LargeMvC-Net	[1,9]	[c,3c]	–	–
MLRR	[1e-2,5e2]	[5e-3,1]	[1,5]	–
SMSC-AND	[1,10]	–	–	–
NLRSC	[1e-6,10]	[1e-6,10]	–	–
AMvSC	[1e-5,0.5]	[1e-5,0.5]	[1,10]	–
SLMVGC	[1e-4,1e4]	[1e-4,1e4]	[1e-1,1e2]	[1,1e3]
LSGMC	[0.01,0.9]	[1e-5,1e2]	–	–
T-UMC	[0.3,1.2]	[0.3,1.2]	[0.3,1.2]	–
OLR-MVC	[1e-3,1e3]	[1e-3,1e3]	[60,100]	–

Table 7: List of notations.

Notations	Descriptions
n, V, c	Number of samples, views, and clusters
d_v	Feature dimension of the v-th view
k	Dimension of the factorized matrices
I	Number of factor matrices
$X^{(v)} \in \mathbb{R}^{d_v \times n}$	Data matrix of the v-th view
$Z^{(v)} \in \mathbb{R}^{n \times n}$	Self-representation matrix of the v-th view
$Z_i^{(v)}$	The i-th factor matrix of $Z^{(v)}$
$E^{(v)}$	Error matrix of the v-th view
$W_i^{(v)}$	Learnable weight matrix of the i-th factor
$\rho_i^{(v)}$	Learnable step-control parameter
p_i	Schatten parameter of the i-th factor
λ_1, λ_2	Weights of the error and reconstruction terms
θ, τ	Weights of the consistency and symmetry terms
α	Parameter for affinity construction