Multi-view Unsupervised Feature Selection with Unified Measurement of Consistency and Diversity

Shengke Xu, Xijiong Xie, Guoqing Chao, Yujie Xiong

PII: S0031-3203(25)01391-3
DOI: https://doi.org/10.1016/j.patcog.2025.112728
Reference: PR 112728

To appear in: *Pattern Recognition*

Received date: 31 August 2024
Revised date: 7 November 2025
Accepted date: 8 November 2025

Please cite this article as: Shengke Xu, Xijiong Xie, Guoqing Chao, Yujie Xiong, Multi-view Unsupervised Feature Selection with Unified Measurement of Consistency and Diversity, *Pattern Recognition* (2025), doi: https://doi.org/10.1016/j.patcog.2025.112728

Highlights

- Fully consider the global and local geometric structures of multi-view data
- Introducing tensors to capture high-order correlations between different views
- Learning the latent representation for unsupervised feature selection
- An efficient solver is developed to solve the proposed optimization problem

Multi-view Unsupervised Feature Selection with Unified Measurement of Consistency and Diversity

Shengke Xu[a], Xijiong Xie[a,b,*], Guoqing Chao[c], Yujie Xiong[d]

[a]*School of Information Science and Engineering, Ningbo University, Ningbo 315211, China*
[b]*Key Laboratory of Mobile Network Application Technology of Zhejiang Province, Ningbo 315211, China*
[c]*School of Computer Sciences and Technology, Harbin Institute of Technology, Weihai, Shandong, 264209, China*
[d]*School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, PR China*

Abstract

The low cost and high efficiency of multi-view unsupervised feature selection (MvUFS) have greatly stimulated research interest in this field. However, existing graph-based MvUFS methods typically focus solely on either the consistency or the diversity of multi-view data, let alone jointly considering both. Moreover, most approaches rely on matrix optimization while neglecting the exploration of higher-order correlations. In this work, we propose a novel multi-view learning framework for unsupervised feature selection to address these problems. First, a unified module is designed to jointly measure consistency and diversity, enabling the construction of a pure graph for each view. These pure graphs are then fused to generate a consensus graph, which, together with latent representations, mutually constrains and facilitates the learning of optimal pure graphs. Furthermore, we employ a low-rank tensor to preserve high-order correlations among views. The proposed methods are seamlessly integrated into a unified framework. Extensive experiments demonstrate that our model outperforms several state-of-the-art feature selection algorithms.

Keywords: Unified measurement, Tensor analysis, Multi-view learning, Latent representation.

*Corresponding author at: School of Information Science and Engineering, Ningbo University, Ningbo 315211, China.

Email addresses: 2211100288@nbu.edu.cn (Shengke Xu), xjxie11@gmail.com (Xijiong Xie), guoqingchao@hit.edu.cn (Guoqing Chao), xiong@sues.edu.com (Yujie Xiong)

1. Introduction

Nowadays, the rapid advancement of imaging and sensing technologies has significantly broadened the application of multi-view data [1]. Multi-view data refers to information collected from multiple perspectives or using different devices. For example, in medical imaging, various techniques such as CT, MRI, and X-ray [2] provide complementary views of the same subject; in biochemistry, drugs and proteins can be represented by their structural and chemical views [3]; and in autonomous driving, multi-view data is obtained by integrating information from on-board cameras, LiDAR, and radar sensors [4], which enables vehicles to perceive their environment with the greater accuracy. The study of multi-view data processing has garnered significant attention [5, 6]. However, despite the richer information and more comprehensive analytical capabilities offered by multi-view data compared to single-view data, challenges arise, such as increased computational complexity and storage requirements, redundant information across views, and the heterogeneity between views, which collectively complicate data fusion [7, 8, 9].

To address the aforementioned challenges, dimensionality reduction algorithms have emerged as a prominent research focus. These algorithms aim to retain the most informative features in the original data while eliminating noise and redundant features. In machine learning, the two most representative dimensionality reduction methods are feature extraction and feature selection, respectively [10, 11]. Feature extraction involves transforming high-dimensional data into a lower-dimensional space, while feature selection identifies a subset of original features, effectively preserving the spatial structure of the original data and reducing the risk of overfitting. This paper focuses on exploring methods for multi-view feature selection. Multi-view feature selection can be categorized based on the availability of sample labels: supervised [12], semi-supervised [13, 14, 15], and unsupervised [11, 16, 17, 18]. In the context of big data, obtaining labeled data is often costly, making unsupervised multi-view feature selection particularly relevant and practically significant.

Current graph-based feature selection methods have demonstrated their effectiveness in various applications. However, several limitations persist. First of all, there

are noise and outliers in real-life data sets, which will cause the subsequently selected features not to have the best discriminant information. Secondly some methods simply consider the consistency [19] and diversity [20] of multi-view data independently, neglecting the underlying relationship between the two. Overemphasizing consistency can lead to the loss of distinctive, discriminative information within individual view, while an excessive focus on diversity weakens the consensus representation across views. Although certain approaches account for both consistency and diversity [21], they treat consistency and diversity as separate modules, failing to fully capture their intrinsic connections. Additionally, most methods employ matrix optimization strategies, which often overlook inter-view correlations and struggle to explore higher-order relationships that ensure global consistency. This limitation hampers the effective integration of fundamental information from multiple views.

In our approach, we fully address the aforementioned issues and propose Multi-view Unsupervised Feature Selection with Unified Measurement of Consistency and Diversity (UCDMvFS). This model can explore consistency and diversity at the same time, extract complementary information between multiple views, and the inherent consistent pure part and relative diversity part of each view by detecting and removing extremely special diversity part in the views. Firstly, our method uses graph learning to generate a similarity matrix for each view, which is then detected and compared with a pre-trained high-quality graph. Accordingly, the inherent consistent pure part and relative diversity part of each view are extracted and fused into a consensus structure graph with precisely required connected parts. Secondly, the consensus structure graph is processed by symmetric non-negative matrix factorization, where latent representations are used to mutually constrain and promote the learning of optimal pure graphs. Finally, multiple pure graphs are stacked into a third-order tensor with low-rank constraints to reserve high-order correlations between views [22, 23]. UCDMvFS integrates unified measurement, tensor analysis, pure graph learning, latent representations, and feature selection into a unified optimization model. To provide a more intuitive representation of our proposed method, we visualize UCDMvFS in Fig. 1. The main contributions of this work are as follows.

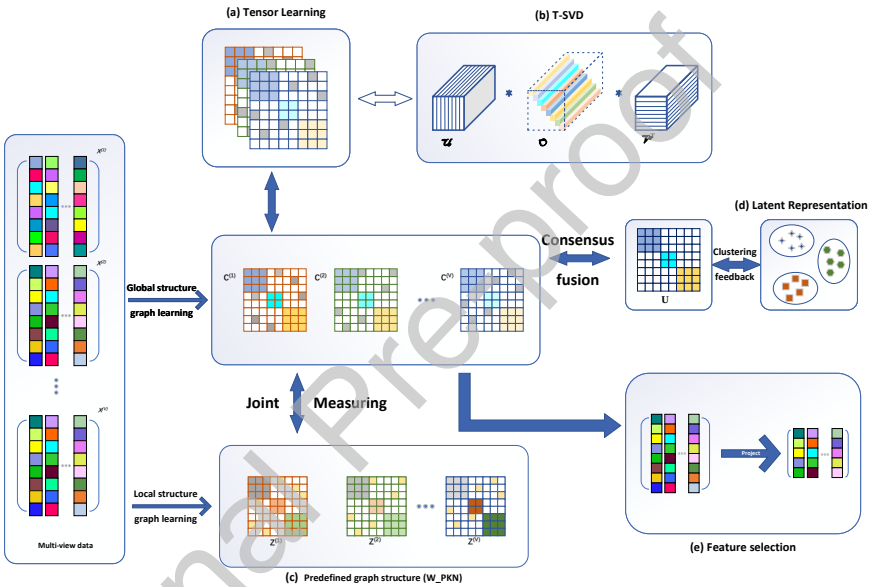


Fig 1: The framework of the proposed UCDMvFS method is provided. The method mainly includes the following modules: (a) stacking pure graphs into third-order tensors to learn high-order information; (b) applying T-SVD-based tensor nuclear norm to explore high-order correlations between multiple views; (c) pre-trained high-quality graphs; (d) clustering feedback on the consistent structure graph using latent representation; (e) the final feature selection process.

1. We propose a new method to uniformly measure the consistency and diversity information of views, which can promote mutual learning during the optimization process.
2. We use a third-order tensor with low-rank constraints to preserve the high-order correlation between views and retain more complete global structural information.
3. We develop an effective optimization algorithm to solve the optimization problem of the objective function and demonstrate the effectiveness of the proposed method through a large number of comprehensive clustering experiments.

The structure of this paper and an overview of its subsequent sections are outlined as follows. Section 2 briefly introduces algorithms closely related to our approach. In Section 3, we provide a detailed description of the proposed method, including the optimization process and model analysis. Section 4 details the experimental setup and presents the results. Finally, in Section 5, we present the conclusions of this study and outline potential future research directions.

2. Related Works

2.1. Notations

First, we elucidate the notations employed in this paper. In our notation, tensors are denoted by uppercase script letters, matrices are represented by boldface capital letters, vectors by bold lowercase letters, and scalars by italic letters. Given an arbitrary matrix $\mathbf{X} \in \mathbb{R}^{d\times n}$, X_{ij} denotes its (i, j)-th entry, while $\mathbf{x}^i$ and $\mathbf{x}_j$ denote its i-th row and j-th column, respectively. Table 1 lists all the commonly used symbols in the paper. Next, we briefly explain the norms employed in this paper: $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{d} X_{ij}^2}$ is the Frobenius norm of $\mathbf{X}$, the $\ell_{2,1}$-norm is defined as $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{n}\sqrt{\sum_{j=1}^{d} X_{ij}^2}$, and the tensor nuclear norm based on t-SVD is defined as $\|\mathcal{C}\|_{\circledast} = \left\|(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1})\,\mathrm{bcicr}(\mathcal{C})\left(\mathbf{F}_{n_3}^{*} \otimes \mathbf{I}_{n_2}\right)\right\|_{*}$.

2.2. Consistency and diversity learning

Significant progress has been made in mining consistency and diversity information when dealing with multi-view data. In multi-view clustering algorithms, some

Table 1: Notations.

Symbols	Definition and description
n	The number of data instances
d^v	The feature dimension of the v-th view
V	The number of views
$\mathbf{X}^v \in \mathbb{R}^{d^v \times n}$	The data matrix of the v-th view
$\mathbf{x}_i^v \in \mathbb{R}^{d^v \times 1}$	The i-th row of the matrix $\mathbf{X}^v$
$\mathbf{C_v}$	The similarity matrix of the v-th view
$\mathbf{H}^v \in \mathbb{R}^{d^v \times c}$	The feature selection matrix of the v-th view
$\mathbf{U} \in \mathbb{R}^{n \times n}$	The consistency representation matrix
$C \in \mathbb{R}^{n_1 \times n_2 \times n_3}$	The 3-order tensor
$\mathbf{F}_{n_3} \in \mathbb{R}^{n_3 \times n_3}$	The Discrete Fourier Transform matrix
$\mathbf{Z}^v \in \mathbb{R}^{n \times n}$	The predefined similarity matrix of the v-th view
$\mathbf{I_c}$, $\mathbf{I_n}$	The identity matrix
$\alpha, \beta, \eta, \lambda$	The hyperparameters
$\mathbf{X}^T$	The transpose operator of matrix $\mathbf{X}$
$\otimes$	The Kronecker product
Tr$(\cdot)$	The trace operator of a matrix
diag$(\cdot)$	The diagonal elements of a matrix
bcirc$(\cdot)$	The operation of a block-cyclic matrix

approaches focus solely on either consistency [24, 25] or diversity [26, 27], while others consider both to capture more comprehensive information. For instance, Li et al. [28] proposed a method that simultaneously accounts for diversity and consistency in both the data space and the learned label space, aiming to learn a pure and robust label matrix for multi-view clustering tasks. Huang et al. [29] introduced a method that jointly measures consistency and diversity, enabling these complementary criteria to be seamlessly integrated into the overall design of the clustering algorithm. Zhang et al. [30] proposed a separable consistency and diversity feature learning method to address the conflict between consistency alignment and reconstruction objectives. Hao et al. [31] incorporated a graph regularizer into low-rank tensor representation learning to uncover the consistent manifold information embedded within multi-view data. Mi et al. [32] proposed learning both a consistent representation and a diverse set of rep-

resentations within a latent embedding space, leading to the learning of an improved affinity matrix.

In multi-view feature selection algorithms, consistency and diversity are also considered. For example, Cao et al. [8] simultaneously accounted for graph heterogeneity and label consistency, effectively exploring both heterogeneous and homogeneous information to enhance feature selection tasks. Tang et al. [21] leveraged shared and partitioned information between different views by projecting it into a label space composed of both consensus and view-specific parts. Cao et al. [33] generated multiple mutually exclusive graphs from views to enhance the complementary information across views. Huang et al. [34] proposed a similarity graph reconstruction model guided by complementary learning, where sparse linear combinations of multiple similarity matrices derived from different views were used to obtain a complete similarity graph for each view. However, while these methods have yielded satisfactory results, a unified measurement of consistency and diversity has not yet been developed for feature selection.

2.3. Structural graph learning

In unsupervised feature selection, applying graph structures has significantly improved the algorithm accuracy. Early methods employed spectral analysis to preserve the data's local geometric structure [35]. Local structure emphasizes the data relationship within the local neighborhoods. This approach is inspired by traditional dimensionality reduction methods such as Laplacian Eigenmaps (LE) and its linear extension, Locality Preserving Projections (LPP). Similarly, subsequent methods introduced graph regularization to retain the local geometric structure [16]. Specifically, the points that are adjacent in high-dimensional space should remain adjacent in low-dimensional space. This can be expressed as follows.

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\mathbf{W}^{vT}\mathbf{x}_i^v-\mathbf{W}^{vT}\mathbf{x}_j^v\right\|^2 S_{ij}^v+\Omega(\mathbf{W}), \tag{1}$$

where S_{ij}^v represents the local similarity graph for the v-th view, $\mathbf{W}^v$ is the projection matrix for the v-th view, and $\Omega(\mathbf{W})$ denotes the regularization or constraint on $\mathbf{W}$. During the same period, the concept of preserving global structures was introduced, utiliz-

ing subspace learning and self-representation learning to retain the global structure of the data [17, 36]. The global structure emphasizes the data relationship of the overall data. Given that both local and global structures are beneficial for feature selection, some approaches have incorporated and preserved both types of structural information [37, 38]. Wu et al. [39] proposed an innovative method that jointly performs structural learning and feature selection, exploring the intrinsic relationship between the two. Cao et al. [8] have extended this framework to the multi-view data domain, as outlined below.

$$\min_{\mathbf{H}_v \mathbf{S}_v} \sum_{v} \left\| \mathbf{H}_v^T \mathbf{X}_v - \mathbf{H}_v^T \mathbf{X}_v \mathbf{S}_v \right\|_F^2 + \alpha \left\| \mathbf{S}_v \right\|_F^2 + \Omega\left(\mathbf{H}_v\right), \tag{2}$$

where $\mathbf{S}_v \in \mathbb{R}^{n \times n}$ denotes the similarity matrix for each view and $\mathbf{H}_v \in \mathbb{R}^{d \times c}$ denotes the feature weight matrix for each view. The last two terms are regularization terms imposed on $\mathbf{S}_v$ and $\mathbf{H}_v$, respectively.

2.4. Tensor learning

For the tensor learning the first step involves modeling each view of the same dimensionality using either graph-based or self-representation-based schemes. Subsequently, all similarity graphs are stacked into a third-order tensor to facilitate tensor-based operations and constraints. Similar to low-rank representation (LRR) methods, which enforce low-rank properties via the nuclear norm, tensor-based approaches aim to find a tight relaxation of the tensor rank to achieve the same objective. Three mainstream tensor decomposition techniques exist: CANDECOMP/PARAFAC (CP) [40], Tucker [41], and Tensor Singular Value Decomposition (t-SVD) [42]. Among them, the tensor nuclear norm based on t-SVD offers the tightest convex relaxation of the tensor multi-rank [43]. As a result, t-SVD based multi-view feature selection methods generally exhibit the superior performance.

Our survey indicates that Zhang et al. [44] were the first to propose tensor-based multi-view feature selection, effectively exploring high-order correlations among multi-view data. Wang et al. [45] constructed pseudo-labels for each view and applied tensor learning on these labels to capture high-order correlations. Yuan et al. [46] employed tensor learning on local graph structures to enhance the high-order relationships be-

tween views. Liang et al. [47] utilized tensor learning on the denoised portions of the learned similarity graphs to capture high-quality high-order information.

The application of tensor methods in multi-view unsupervised feature selection (MvUFS) remains relatively limited. Our experiments with several method combinations revealed that applying tensor learning to the global structure graph enables more comprehensive learning of high-order correlations between views, effectively enriching the inter-view information.

3. Proposed Method

3.1. Model Framework

The proposed approach employs a multi-view data processing framework that integrates structure learning with feature selection. An affinity graph is independently constructed for each view, and the specific form is given in Eq.(3) as follows.

$$\begin{aligned} \min_{\mathbf{H}_v, \mathbf{C}_v} \sum_{v=1}^{V} \left\| \mathbf{H}_v^T \mathbf{X}_v - \mathbf{H}_v^T \mathbf{X}_v \mathbf{C}_v \right\|_F^2 + \|\mathbf{H}_v\|_{2,1} + \|\mathbf{C}_v\|_F^2 \\ \text{s.t. } \mathbf{H}_v^T \mathbf{X}_v \mathbf{X}_v^T \mathbf{H}_v = \mathbf{I}_c, \end{aligned} \tag{3}$$

where the last two terms represent regularization terms, used to constrain both the feature selection component and the self-representation reconstruction component. Among the additional constraints, the PCA-like constraint $\mathbf{H}_v^T \mathbf{X}_v \mathbf{X}_v^T \mathbf{H}_v = \mathbf{I}_c$ is employed to minimize redundancy in the original data as much as possible.

The aforementioned paradigm effectively captures the global structure within each view but does not explicitly account for consistency and diversity across views. Consistency captures the shared, significant information across multiple views, while diversity helps uncover unique feature information. In multi-view unsupervised feature selection, balancing these two aspects is a key challenge. Overemphasizing consistency may lead to overlooking unique information in individual view, whereas excessive emphasis on diversity could result in neglecting commonalities across views. Therefore, designing algorithms and strategies that appropriately balance these factors maintaining alignment across different views while fully leveraging their differences is crucial

for better feature selection in multi-view data. Consequently, we propose introducing consistency and diversity learning into the model.

The objective of consistency is to learn a consensus graph $\mathbf{U}$. We can obtain consistency information from similarity graphs by integrating multiple similarity graphs $\{\mathbf{C_1}, \mathbf{C_2}, \cdots, \mathbf{C_v}\}$. This can be formulated as

$$\begin{aligned} \min_{\mathbf{U}} \quad & \textstyle\sum_{v=1}^{V} \|\mathbf{U} - \mathbf{C}_v\|_F \\ \text{s.t.} \quad & \mathbf{u}^T\mathbf{1} = 1, \mathbf{U} \geq 0, \end{aligned} \tag{4}$$

where $\mathbf{u}$ represents a column vector of $\mathbf{U}$. This equation does not incorporate view-specific weights. To avoid introducing additional hyperparameters through the weights, we adopt an inverse distance weighting strategy [48], in which the weight of the i-th view is defined as

$$\mu^v = \frac{1}{2\|\mathbf{U} - \mathbf{C}_v\|_F}, \tag{5}$$

where the weight μ^v is entirely data-driven. Incorporating the self-learned weight μ^v, Eq.(4) can be reformulated as

$$\begin{aligned} \min_{\mathbf{U}} \quad & \textstyle\sum_{v=1}^{V} \mu^v \|\mathbf{U} - \mathbf{C}_v\|_F^2 \\ \text{s.t.} \quad & \mathbf{u}^T\mathbf{1} = 1, \mathbf{U} \geq 0. \end{aligned} \tag{6}$$

In terms of diversity, we consider the potential discrepancies between the local and global structures, both of which are crucial for feature selection. Therefore, we employ the *W_PKN* algorithm [49] to pre-construct high-quality local structure information for each view, denoted as $\{\mathbf{Z_1}, \mathbf{Z_2}, \cdots, \mathbf{Z_v}\}$. Next, we compute the difference between the local structure graph and the global structure graph obtained from Eq.(3), denoted as $\mathbf{Z}_i - \mathbf{C}_i$. If the diversity between views is sparse, the sum of the products across views should be minimized. At the same time, we aim to minimize intra-view divergence [26]. Based on these considerations, the expression can be formulated as follows.

$$\begin{aligned} \min_{\mathbf{C}_v} \quad & \sum_{v,w=1}^{V} \mathfrak{w}_{vw}\mu^v\mu^w \operatorname{Tr}\left((\mathbf{Z}_v - \mathbf{C}_v)(\mathbf{Z}_w - \mathbf{C}_w)^T\right) \\ \text{s.t.} \quad & \mathbf{Z}_v \geq \mathbf{C}_v \geq 0, v, w = \{1, \ldots, V\}, \end{aligned} \tag{7}$$

where $\mathfrak{W} = [\mathfrak{w}_{vw}] \in \mathbb{R}^{V\times V}$ is a square matrix, with the off-diagonal elements represented by λ and the diagonal elements by η, respectively. By combining Eqs.(3), (6),

and (7), the rewritten expression is as follows.

$$\begin{aligned}
&\min_{\mathbf{H}_v,\mathbf{C}_v,\mathbf{U}} \sum_{v=1}^{V} \left\|\mathbf{H}_v^T\mathbf{X}_v - \mathbf{H}_v^T\mathbf{X}_v\mathbf{C}_v\right\|_F^2 + \sum_{v=1}^{V} \|\mathbf{H}_v\|_{2,1} \\
&+ \|\mathbf{C}_v\|_F^2 + \sum_{v=1}^{V} \mu^v \|\mathbf{U} - \mathbf{C}_v\|_F^2 \\
&+ \sum_{v,w=1}^{V} \varpi_{vw}\mu^v\mu^w \operatorname{Tr}\left((\mathbf{Z}_v - \mathbf{C}_v)(\mathbf{Z}_w - \mathbf{C}_w)^T\right) \\
&\text{s.t.} \quad \mathbf{H}_v^T\mathbf{X}_v\mathbf{X}_v^T\mathbf{H}_v = \mathbf{I}_c, \mathbf{u}^T\mathbf{1} = 1, \mathbf{U} \geq 0, \\
&\qquad \mathbf{Z}_v \geq \mathbf{C}_v \geq 0, v, w = \{1, \ldots, V\}.
\end{aligned} \tag{8}$$

High-order information provides a more comprehensive representation of correlations between data. Therefore, we introduce tensor analysis to the global graph structure by stacking $\{\mathbf{C_1}, \mathbf{C_2}, \cdots, \mathbf{C_v}\}$ into a third-order tensor $\mathcal{C} \in \mathbb{R}^{n\times n\times V}$, while ensuring the low-rank property of $\mathcal{C}$ through the tensor nuclear norm based on t-SVD. Consequently, Eq.(8) can be reformulated as follows.

$$\begin{aligned}
&\min_{\mathbf{H}_v,\mathbf{C}_v,\mathbf{U},\mathcal{C}} \sum_{v=1}^{V} \left\|\mathbf{H}_v^T\mathbf{X}_v - \mathbf{H}_v^T\mathbf{X}_v\mathbf{C}_v\right\|_F^2 + \sum_{v=1}^{V} \|\mathbf{H}_v\|_{2,1} \\
&+ \|\mathcal{C}\|_{\circledast} + \sum_{v=1}^{V} \mu^v \|\mathbf{U} - \mathbf{C}_v\|_F^2 \\
&+ \sum_{v,w=1}^{V} \varpi_{vw}\mu^v\mu^w \operatorname{Tr}\left((\mathbf{Z}_v - \mathbf{C}_v)(\mathbf{Z}_w - \mathbf{C}_w)^T\right) \\
&\text{s.t.} \quad \mathbf{H}_v^T\mathbf{X}_v\mathbf{X}_v^T\mathbf{H}_v = \mathbf{I}_c, \mathbf{u}^T\mathbf{1} = 1, \mathbf{U} \geq 0, \\
&\qquad \mathbf{Z}_v \geq \mathbf{C}_v \geq 0, v, w = \{1, \ldots, V\},
\end{aligned} \tag{9}$$

where $\mathcal{C} = \Phi(\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_v)$, with $\mathbf{C}_v \geq 0$. $\Phi(\cdot)$ denotes the operation of forming the third-order tensor $\mathcal{C}$ by stacking the global structure graphs from different views.

To enhance interconnections among data, we decompose the data into a latent space, extracting latent relational information between them. Specifically, the latent representations of different samples interact to form connection information, where samples with similar latent representations are more likely to be connected compared to those with distinct latent representations. Typically, the latent representation of connection information is formed through a symmetric non-negative matrix factorization

model, which decomposes the consensus similarity matrix $\mathbf{U}$ into the product of a non-negative matrix $\mathbf{V}$ and its transpose $\mathbf{V}^T$ in a lower-dimensional latent space. By incorporating this into Eq.(9), our model can be ultimately expressed as follows.

$$\begin{aligned}
&\min_{\mathbf{H}_v,\mathbf{C}_v,\mathbf{U},C,\mathbf{V}} \sum_{v=1}^{V} \left\|\mathbf{H}_v^T\mathbf{X}_v - \mathbf{H}_v^T\mathbf{X}_v\mathbf{C}_v\right\|_F^2 + \alpha \|C\|_{\circledast} + \beta \sum_{v=1}^{V} \|\mathbf{H}_v\|_{2,1} \\
&+ \sum_{v,w=1}^{V} \varpi_{vw}\mu^v\mu^w \operatorname{Tr}\left((\mathbf{Z}_v - \mathbf{C}_v)(\mathbf{Z}_w - \mathbf{C}_w)^T\right) \\
&+ \sum_{v=1}^{V} \mu^v \|\mathbf{U} - \mathbf{C}_v\|_F^2 + \left\|\mathbf{U} - \mathbf{V}\mathbf{V}^T\right\|_F^2 \qquad (10) \\
&\text{s.t.} \quad \mathbf{H}_v^T\mathbf{X}_v\mathbf{X}_v^T\mathbf{H}_v = \mathbf{I}_c, \mathbf{u}^T\mathbf{1} = 1, \mathbf{U} \geq 0, \mathbf{V} \geq 0, \\
&\qquad \mathbf{V}^T\mathbf{V} = \mathbf{I}_c, \mathbf{Z}_v \geq \mathbf{C}_v \geq 0, v, w = \{1, \ldots, V\},
\end{aligned}$$

where $\mathbf{V}$ is the latent representation matrix, which can be used as a clustering structure for the data to reduce the negative impact of noisy connections in the affinity graph, thereby improving the overall robustness of the model, while α and β are balance hyperparameters.

3.2. Optimization

Due to the interdependence of variables in Eq.(10), direct optimization is challenging. To overcome this difficulty, we developed an alternating optimization strategy that iteratively optimizes one variable while keeping the others fixed.

First, we introduce an auxiliary tensor $\mathcal{G}$ to decouple the variable updates. As a result, Eq.(10) is reformulated as follows.

$$\begin{aligned}
&\min_{\mathbf{H}_v,\mathbf{C}_v,\mathbf{U},\mathcal{G},\mathbf{V}} \sum_{v=1}^{V} \left\|\mathbf{H}_v^T\mathbf{X}_v - \mathbf{H}_v^T\mathbf{X}_v\mathbf{C}_v\right\|_F^2 + \alpha \|\mathcal{G}\|_{\circledast} + \beta \sum_{v=1}^{V} \|\mathbf{H}_v\|_{2,1} \\
&+ \sum_{v,w=1}^{V} \varpi_{vw}\mu^v\mu^w \operatorname{Tr}\left((\mathbf{Z}_v - \mathbf{C}_v)(\mathbf{Z}_w - \mathbf{C}_w)^T\right) \\
&+ \sum_{v=1}^{V} \mu^v \|\mathbf{U} - \mathbf{C}_v\|_F^2 + \left\|\mathbf{U} - \mathbf{V}\mathbf{V}^T\right\|_F^2 \qquad (11) \\
&\text{s.t.} \quad \mathbf{H}_v^T\mathbf{X}_v\mathbf{X}_v^T\mathbf{H}_v = \mathbf{I}_c, \mathbf{u}^T\mathbf{1} = 1, \mathbf{U} \geq 0, \mathbf{V} \geq 0, \\
&\qquad \mathbf{V}^T\mathbf{V} = \mathbf{I}_c, \mathbf{Z}_v \geq \mathbf{C}_v \geq 0, v, w = \{1, \ldots, V\}, \mathcal{C} = \mathcal{G}.
\end{aligned}$$

Ultimately, we optimize the proposed method using an augmented Lagrange multiplier-based alternating direction method of multipliers (ALM-ADMM) [50]. The final form of the augmented Lagrangian function is expressed as follows.

$$
\begin{aligned}
&\mathcal{L}(\mathbf{H}_v, \mathbf{C}_v, \mathcal{G}, \mathbf{U}, \mathbf{V}) = \\
&\min_{\mathbf{H}_v, \mathbf{C}_v, \mathbf{U}, \mathcal{G}, \mathbf{V}} \sum_{v=1}^{V} \left\| \mathbf{H}_v^T \mathbf{X}_v - \mathbf{H}_v^T \mathbf{X}_v \mathbf{C}_v \right\|_F^2 + \alpha \|\mathcal{G}\|_{\circledast} + \beta \sum_{v=1}^{V} \|\mathbf{H}_v\|_{2,1} \\
&+ \sum_{v,w=1}^{V} \mathfrak{w}_{vw} \mu^v \mu^w \operatorname{Tr}\left((\mathbf{Z}_v - \mathbf{C}_v) (\mathbf{Z}_w - \mathbf{C}_w)^T \right) \\
&+ \sum_{v=1}^{V} \mu^v \|\mathbf{U} - \mathbf{C}_v\|_F^2 + \left\| \mathbf{U} - \mathbf{V}\mathbf{V}^T \right\|_F^2 \\
&+ \frac{\rho}{2} \left\| \mathcal{G} - C - \frac{\mathcal{P}}{\rho} \right\|_F^2 \\
&\text{s.t.} \quad \mathbf{H}_v^T \mathbf{X}_v \mathbf{X}_v^T \mathbf{H}_v = \mathbf{I}_c, \mathbf{u}^T \mathbf{1} = 1, \mathbf{U} \geq 0, \mathbf{V} \geq 0, \\
&\qquad \mathbf{V}^T \mathbf{V} = \mathbf{I}_c, \mathbf{Z}_v \geq \mathbf{C}_v \geq 0, v, w = \{1, \ldots, V\},
\end{aligned}
\tag{12}
$$

where $\mathcal{P}$ represents the Lagrange multipliers and $\rho > 0$ denotes the penalty parameter. We employ an efficient alternating optimization strategy to iteratively solve Eq.(12). ADMM effectively decomposes the problem into multiple subproblems, each of which is solved independently, thereby reducing the overall complexity of the optimization. The detailed alternating updated scheme is as follows.

$\mathcal{G}-$**subproblem** :By fixing the other variables, the terms related to $\mathcal{G}$ in Eq.(12) can be extracted and expressed as follows.

$$
\min_{\mathcal{G}} \quad \alpha \|\mathcal{G}\|_{\circledast} + \frac{\rho}{2} \left\| \mathcal{G} - C - \frac{\mathcal{P}}{\rho} \right\|_F^2, \tag{13}
$$

Based on the tensor nuclear norm minimization problem outlined in [51], the optimization of $\mathcal{G}$ can be achieved through the following steps.

$$
\mathcal{G}^* = \mathcal{R}_{\rho'} \left(C + \frac{1}{\rho} \mathcal{P} \right) = \mathcal{U} * \mathcal{R}_{\rho'}(\mathcal{O}) * \mathcal{V}^{\mathrm{T}}, \tag{14}
$$

where $\rho' = n/\rho$, $\left(C + \frac{1}{\rho}\mathcal{P}\right) = \mathcal{U} * \mathcal{O} * \mathcal{V}^{\mathrm{T}}$ and $\mathcal{R}_{\rho'}(\mathcal{O}) = \mathcal{O} * \mathcal{J}$. Here, $\mathcal{J}$ is the f-diagonal tensor, with its diagonal elements in the Fourier domain represented as $\mathcal{J}_f(i,i,j) = \max\left(0, 1 - \frac{\rho'}{\mathcal{O}_f^{(j)}(i,i)}\right)$.

$(\mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_V)-$**subproblem** :By removing irrelevant items, we need to solve the Eq.(12) which is simplified as the following form.

$$\begin{aligned}
&\min_{\mathbf{C}_v} \sum_{v=1}^{V} \left\| \mathbf{H}_v^T \mathbf{X}_v - \mathbf{H}_v^T \mathbf{X}_v \mathbf{C}_v \right\|_F^2 + \sum_{v=1}^{V} \mu^v \left\| \mathbf{U} - \mathbf{C}_v \right\|_F^2 \\
&+ \sum_{v,w=1}^{V} \varpi_{vw} \mu^v \mu^w \operatorname{Tr}\left((\mathbf{Z}_v - \mathbf{C}_v) (\mathbf{Z}_w - \mathbf{C}_w)^T \right) \\
&+ \frac{\rho}{2} \left\| \mathbf{G}_v - \mathbf{C}_v - \frac{\mathbf{P}_v}{\rho} \right\|_F^2, \\
&\text{s.t.} \quad \mathbf{Z}_v \geq \mathbf{C}_v \geq 0.
\end{aligned} \tag{15}$$

By calculating the derivative of Eq.(15) with respect to $\mathbf{C}_v$ and letting it to be zero, we can get the updating rule of $\mathbf{C}_v$ as below.

$$\begin{aligned}
\mathbf{C}_v &= \max\left(\frac{2\mathbf{K}_v + \xi \mathbf{E}_v + 2\mu^v \mathbf{U} + \rho \mathbf{G}_v - \mathbf{P}_v}{2\mathbf{K}_v + 2\mu^v \mathbf{I} + \rho \mathbf{I}}, 0 \right), \\
\mathbf{C}_v &= \min(\mathbf{Z}_v, \mathbf{C}_v),
\end{aligned} \tag{16}$$

where $\xi = \sum_{v,w}^{V} \varpi_{vw} \mu^v \mu^w$, $\mathbf{E}_v = \mathbf{Z}_v - \mathbf{C}_v$ and $\mathbf{K}_v = \mathbf{X}_v^T \mathbf{H}_v \mathbf{H}_v^T \mathbf{X}_v$.

$\mathbf{H}_v-$**subproblem** :Extracting the relevant terms, the subproblem concerning $\mathbf{H}_v$ is formulated as follows.

$$\begin{aligned}
&\min_{\mathbf{H}_v} \left\| \mathbf{H}_v^T \mathbf{X}_v - \mathbf{H}_v^T \mathbf{X}_v \mathbf{C}_v \right\|_F^2 + \beta \left\| \mathbf{H}_v \right\|_{2,1} \\
&\text{s.t. } \mathbf{H}_v^T \mathbf{X}_v \mathbf{X}_v^T \mathbf{H}_v = \mathbf{I}_c.
\end{aligned} \tag{17}$$

The presence of the $\ell_{2,1}$-norm necessitates the introduction of a diagonal matrix $\mathbf{D}_v$, with its elements obtained through the following equation.

$$\mathbf{D}_{v[i,i]} = \frac{1}{\max\left(2 \left\| \mathbf{H}_{v[i,:]} \right\|_2, \varepsilon \right)}, \tag{18}$$

where ε is introduced to handle the case when $\left\| \mathbf{H}_{v[i,:]} \right\|_2 = 0$. Using $\mathbf{R}_v = (\mathbf{I}_n - \mathbf{C}_v)(\mathbf{I}_n - \mathbf{C}_v)^T$ and Eq.(18), the optimization problem for H is transformed into

$$\begin{aligned}
&\min_{\mathbf{H}_v} \operatorname{Tr}\left[\mathbf{H}_v^T \left(\mathbf{X}_v \mathbf{R}_v \mathbf{X}_v^T + \beta \mathbf{D}_v \right) \mathbf{H}_v \right] \\
&\text{s.t. } \mathbf{H}_v^T \mathbf{X}_v \mathbf{X}_v^T \mathbf{H}_v = \mathbf{I}_c.
\end{aligned} \tag{19}$$

The optimal $\mathbf{H}_v$ can be obtained by solving the following generalized eigen-problem [33].

$$\left(\mathbf{X}_v\mathbf{R}_v\mathbf{X}_v^T + \beta\mathbf{D}_v\right)\mathbf{H}_v = \Gamma_v\mathbf{X}_v\mathbf{X}_v^T\mathbf{H}_v, \tag{20}$$

where Γ_v is a diagonal matrix with its diagonals filled by eigenvalues. However, solving Eq.(20) requires that $\mathbf{X}_v\mathbf{X}_v^T$ be nonsingular. Additionally, the computational complexity of this solution is $O(d^3 + nd^3)$, which is impractical for high-dimensional data. Referring to [33], we obtain the optimal $\mathbf{H}_v$ by solving the following problem.

$$\min_{\mathbf{H}_v} \left\|\mathbf{Y}_v - \mathbf{X}_v^T\mathbf{H}_v\right\|_F^2 + \alpha \left\|\mathbf{H}_v\right\|_{2,1}, \tag{21}$$

where $\mathbf{Y}_v$ consists of eigenvectors corresponding to the c smallest eigenvalues by solving the eigen-problem $\mathbf{R}_v\mathbf{Y}_v = \Gamma_v\mathbf{Y}_v$. Using the diagonal matrix defined in Eq.(18) and the Iterative Reweighted Least-Squares (IRLS) [52] algorithm, the optimal $\mathbf{H}_v$ can be obtained by

$$\mathbf{H}_v = \left(\mathbf{X}_v\mathbf{X}_v^T + \beta\mathbf{D}_v\right)^{-1}\mathbf{X}_v\mathbf{Y}_v. \tag{22}$$

U−subproblem :By fixing the remaining variables, the terms related to $\mathbf{U}$ are extracted as follows.

$$\begin{aligned} &\sum_{v=1}^{V} \mu^v \left\|\mathbf{U} - \mathbf{C}_v\right\|_F^2 + \left\|\mathbf{U} - \mathbf{V}\mathbf{V}^T\right\|_F^2 \\ &\text{s.t.} \quad \mathbf{u}^T\mathbf{1} = 1, \mathbf{U} \geq 0, \mathbf{V} \geq 0, \end{aligned} \tag{23}$$

Evidently, the update for each view is independent, and we express the element-wise form of Eq.(23) as follows.

$$\begin{aligned} &\min_{\mathbf{u}_i} \sum_{i=1}^{m}\sum_{j=1}^{n} \mu^v \left(u_{ij} - c_{ij}^v\right)^2 + \left(u_{ij} - v_i v_j^T\right)^2 \\ &\text{s.t. } \mathbf{u}_t^T\mathbf{1} = 1, u_{ij} \geq 0, \end{aligned} \tag{24}$$

where u_{ij} represents the j-th element corresponding to the row vector $\mathbf{u}_i$. Let $\mathbf{b}_i = (\mathbf{V}\mathbf{V}^T)_i$. After merging, Eq.(24) can be reformulated into a more intuitive expression.

$$\begin{aligned} &\min_{\mathbf{u}_i} \left[\mathbf{u}_i - \frac{\sum_{v=1}^{V}\mu^v\mathbf{c}_i^v + \mathbf{b}_i}{\sum_{v=1}^{V}\mu^v + 1}\right]^2 \\ &\text{s.t. } \mathbf{u}_t^T\mathbf{1} = 1, u_{ij} \geq 0, \end{aligned} \tag{25}$$

Finally, we solve Eq.(25) using the method proposed in [29].

V−subproblem :Through simple algebraic manipulation, the update for $\mathbf{V}$ has the following equivalent formulation.

$$\min_{\mathbf{V}} \operatorname{Tr}\left[\mathbf{V}^T\left(\mathbf{I}_n - 2\mathbf{U}\right)\mathbf{V}\right] \quad \text{s.t. } \mathbf{V}^T\mathbf{V} = \mathbf{I}_c. \tag{26}$$

To clearly and intuitively present the solution process, we summarize the above steps in Algorithm 1.

Algorithm 1 The detailed iteration of the UCDMvFS

Input : Multi-view data set: $\mathbf{X} = \{\mathbf{X}_1, \cdots, \mathbf{X}_V\} \in \mathbb{R}^{d_v \times n}$; the number of clusters classes c; the hyperparameters $\alpha, \beta, \eta, \lambda$.

Initialize : Let $\mathbf{H}_v$, $\mathbf{C}_v$, $\mathbf{V}$ as rand matrix and $\mu_v = 1/V$; $\mathbf{Z}_v$ is initialized by W_PKN; $\mathcal{P} = \mathcal{G} = 0$; set $\rho = 1, \gamma = 2, \rho_{max} = 10^{12}$.

while not converged **do**
  Update $\mathcal{G}$ by solving the problem in Eq.(14);
  Update $\mathbf{C}_v$ by solving the problem in Eq.(16);
  Update $\mathbf{H}_v$ by solving the problem in Eq.(22);
  Update $\mathbf{U}$ by solving the problem in Eq.(25);
  Update $\mathbf{V}$ by solving the problem in Eq.(26);
  Update μ^v by solving the problem in Eq.(5);
  Update $\mathcal{P}$ by $\mathcal{P} = \mathcal{P} + \rho(\mathcal{C} - \mathcal{G})$;
  Update ρ by $\min(\rho\gamma, \rho_{max})$;
End while

Output : $\{\mathbf{H}^v\}_{v=1}^V$;

Feature selection : Compute all $\left\|\mathbf{H}_i^v\right\|_2$ and rank the features in descending order according to their scores, then select the top K features with the highest scores. The final feature subset consists of the corresponding discriminative features.

3.3. Complexity Analysis

Following Algorithm 1, we derive the time complexity of the model through six steps. First, updating the tensor $\mathcal{G}$ involves calculating the tensor FFT, inverse FFT, and the SVD decomposition of an $n \times V$ matrix, with a complexity of $O(n^2V^2 + 2n^2V\log(n))$. Second, the computation of $\mathbf{C}_v$ involves an inversion operation with a complexity of $O(n^3)$. Third, updating the feature weight matrix $\mathbf{H}_v$ requires solving the eigenvalue problem and performing sparse feature selection, with a corresponding computational complexity of $O(cn^2 + d_v^3)$. Fourth, updating $\mathbf{U}$ results in a computational complexity of $O(cn)$. Fifth, updating $\mathbf{V}$ involves SVD, leading to a complexity of $O(cn^2)$. Lastly, the update of μ^v has a complexity of $O(Vn^2)$. Considering that $c \ll n$ and $V \ll n$, the overall computational complexity of the model is $O(n^3 + n^2V^2 + 2n^2V\log(n))$.

3.4. Convergence Analysis

The model consists of five optimization sub-problems, and the overall convergence analysis of the model is difficult. Therefore, we decompose the sub-problems and update one of the variables while keeping the other variables unchanged. For $\mathbf{C}_v$, it is a closed-form solution. For $\mathbf{H}_v$, it includes a $\ell_{2,1}$-norm regularization term, which means that the update $\mathbf{H}_v$ is convex but not smooth. It can be solved using known methods [8, 33]. We use the IRLS algorithm to effectively solve this problem and ensure the convergence of the objective function. For $\mathbf{V}$, based on the Kuhn-Tucker condition [53], the objective function value decreases with iteration, and the optimal solution of $\mathbf{V}$ can be obtained. For $\mathcal{G}$, it satisfies $\|\mathcal{G} - C\|_\infty < \text{eps}$, where eps is a very small value that stops the iteration. In addition, subsection 4.7 will provide experimental evidence to further verify the convergence of the model.

4. Experiments

4.1. Datasets

In our model validation experiments, we evaluated the effectiveness of the UCD-MvFS algorithm using nine different benchmark multi-view datasets. For reference, the detailed information of these multi-view datasets is compiled in Table 2.

Table 2: Detailed information of the multi-view datasets used in our experiments.

Feature	Handwritten	Caltech101-7	MSRCV1	Outdoor scene	ORL	3Sources	yale	WebKB	BBCSport
1	FCCS(76)	GAB(48)	CMT(24)	GIST(512)	View 1(4096)	BBC(3560)	Intensity(4096)	Text(1703)	View 1(3183)
2	KAR(64)	WM(40)	HOG(576)	HOG(432)	View 2(3304)	Reuters(3631)	LBP(3304)	Link(230)	View 2(3203)
3	FAC(216)	CENTRIST(254)	GIST(512)	LBP(256)	View 3(6750)	Guardian(3068)	GABOR(6075)	Title(230)	-
4	PA(240)	HOG(1984)	CENTRIST(254)	GABOR(48)	-	-	-	-	-
5	ZER(47)	GIST(512)	LBP(256)	-	-	-	-	-	-
6	MOR(6)	LBP(928)	-	-	-	-	-	-	-
Instance	2000	1474	210	2688	400	169	165	203	544
Class	10	7	7	8	40	6	15	4	5

4.2. Compared Methods

We used the aforementioned datasets to experimentally compare the proposed method with both classical and state-of-the-art algorithms. The specific details of the algorithms used for comparison are briefly summarized below

- LS [35] is a feature selection method that emphasizes preserving locality by ensuring that distances between similar samples are minimized.
- CGMvUFS [19] learns a latent feature matrix across all views and optimizes the consensus matrix to minimize the dissimilarity between the clustering indicator matrix of each view and the consensus matrix.
- NSGL [54] learns a structured graph directly from raw features by applying hierarchical constraints, simultaneously utilizing the complementary nature of multi-view features for adaptive feature selection.
- TLR [46] generates multiple graphs and aggregates them into a tensor under the low-rank constraint to capture high-order inter-view information.
- TRCA-CGL [47] leverages low-rank tensor learning and consensus graph learning to acquire high-quality local structures and reliable pseudo-cluster labels, which guide feature selection.
- CDMvFS [33] aims to generate multiple mutually exclusive graphs to enhance inter-view complementarity and uses consistency metrics to integrate graph learning and clustering learning.

- JMVFG [55] leverages orthogonal decomposition to obtain clustering indicators for each view, unifying graph learning and MvUFS within a single framework.
- SCMvFS [8] generates a unified clustering indicator through spectral analysis, simultaneously accounting for the heterogeneity of graphs and the consistency of indicators to enhance feature selection.

4.3. Experimental Setup and Metrics

To evaluate the performance differences between the proposed method and the comparison methods more intuitively and effectively, we adjusted several parameter settings based on recommendations from relevant literature on the comparison algorithms. For CGMvUFS, the neighborhood size and bandwidth of the Gaussian kernel function were set to 5 and 1, respectively, with r set to 2. For TLR, the adjustment range was set to $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$ to control consistency. The tensor parameters in our method and TLR were tuned within the range $\{0.0001, 0.001, 0.01, 0.1, 1\}$. For fairness, the remaining parameters were varied within the set $\{0.01, 0.1, 1, 10, 100\}$.

Next, we conducted a series of clustering experiments using the k-means algorithm. Since k-means is sensitive to the initial point selection, we ran it 20 times and calculated the average performance. The number of selected features in the experiments was controlled within the range $\{10, 20, 30, ..., 280, 290, 300\}$. All experiments were conducted in MATLAB R2022b on a machine equipped with an Intel(R) Core(TM) i5-10500 CPU @ 3.10GHz and 16 GB of RAM.

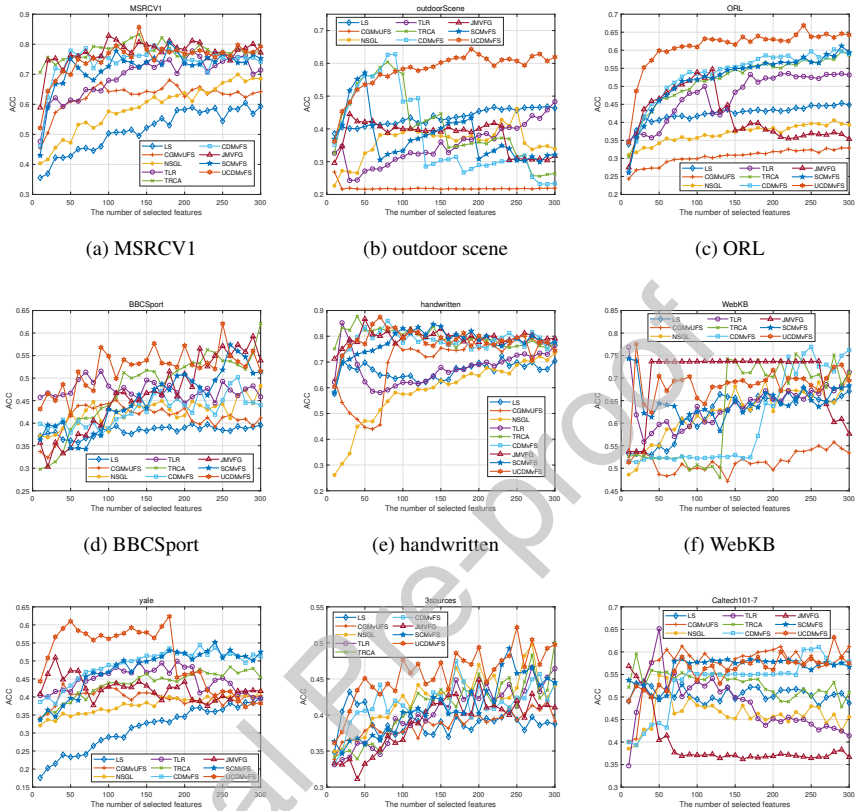
For the evaluation metrics, we selected fours of the most representative indicators: Accuracy (ACC), Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and F-score. Higher values of these metrics indicate better clustering performance of the selected features, suggesting that the features are more suitable for subsequent clustering tasks.

4.4. Experimental Results and Analysis

The results of the four metrics for all methods across different datasets are summarized in Table 3. The best results are highlighted in bold, and the second-best results are underlined. Except for the handwritten dataset, our method consistently achieves either

Table 3: Clustering evaluation metrics across different datasets for various methods.

Datasets	Metrics	LS	CGMvFS	NSGL	TLR	TRCA-CGL	CDMvFS	JMVFG	SCMvFS	UCDMvFS
MSRC_v1	ACC	60.50±5.67	68.07±5.91	69.95±6.12	77.79±9.16	83.29±5.20	79.57±6.44	82.81±6.39	80.50±4.65	**85.69±8.71**
	NMI	48.68±3.52	58.46±3.33	64.26±5.07	72.59±7.46	73.79±2.26	74.15±3.60	77.51±4.65	73.01±4.96	**80.78±6.64**
	ARI	38.36±3.78	48.70±5.15	54.21±5.71	65.52±10.71	69.29±4.45	67.40±2.81	72.00±6.73	66.03±7.84	**75.26±9.68**
	Fscore	47.45±3.08	56.24±4.24	60.82±4.81	70.69±8.92	73.66±3.72	72.14±2.38	76.06±5.68	71.01±6.55	**78.81±8.20**
outdoor_scene	ACC	46.82±2.13	26.85±0.72	46.15±1.62	48.31±2.06	60.42±2.96	62.81±5.14	44.52±3.26	57.12±4.23	**64.32±3.42**
	NMI	40.01±1.04	11.90±0.55	37.33±0.95	37.95±0.85	50.36±0.24	52.01±1.08	36.03±0.98	43.42±0.72	**53.09±0.87**
	ARI	28.21±0.57	6.68±0.34	25.07±1.09	25.48±0.81	40.19±0.41	42.41±1.04	23.57±1.78	34.29±1.46	**43.35±1.17**
	Fscore	37.79±0.48	19.43±0.76	35.04±1.01	35.49±0.71	47.95±0.41	49.88±0.77	34.01±1.59	42.77±1.18	**50.73±0.84**
ORL	ACC	45.30±1.93	32.91±1.39	40.59±1.94	53.55±2.93	59.69±3.70	60.24±3.26	54.71±3.02	61.20±4.59	**66.91±3.43**
	NMI	66.81±1.65	55.29±1.35	62.58±1.02	73.61±1.68	77.36±1.37	78.09±1.73	73.22±1.88	78.59±2.11	**82.17±1.59**
	ARI	28.99±2.07	14.58±1.27	23.67±1.54	39.27±2.98	46.84±2.43	47.80±3.47	39.26±2.90	49.02±4.14	**55.23±3.86**
	Fscore	30.96±1.97	16.95±1.20	25.68±1.49	40.92±2.86	48.21±2.35	49.13±3.36	40.82±2.81	50.31±4.02	**56.36±3.75**
BBCSport	ACC	39.86±2.25	45.12±3.56	48.26±8.61	51.53±6.84	**62.13±8.57**	50.19±8.92	59.26±7.33	57.42±9.94	62.07±8.51
	NMI	9.24±3.69	16.01±4.36	22.52±6.17	31.01±11.91	43.16±10.40	24.49±11.88	**44.43±9.16**	37.07±7.96	43.79±12.25
	ARI	2.77±2.43	8.81±1.78	11.64±7.30	18.47±13.24	**35.58±13.90**	17.50±13.27	29.76±17.01	25.45±13.76	33.56±14.14
	Fscore	39.27±0.09	41.70±1.29	41.96±3.82	46.20±7.13	**56.20±12.58**	45.74±7.04	51.21±10.47	48.96±9.78	54.14±10.98
handwritten	ACC	72.39±7.34	79.17±6.80	73.53±4.73	85.22±5.91	**87.75±5.42**	85.99±8.29	86.71±6.94	84.63±8.04	87.46±5.84
	NMI	69.55±3.14	76.97±3.14	71.34±2.21	82.42±4.38	**85.17±3.01**	83.96±4.83	85.15±3.90	81.84±3.49	83.77±3.58
	ARI	60.49±5.05	70.29±6.05	62.90±3.57	76.25±7.60	**81.20±6.35**	79.31±6.74	81.16±6.85	77.69±6.65	79.78±4.79
	Fscore	64.66±4.47	73.38±5.37	66.74±3.17	78.73±6.75	**83.14±5.63**	81.47±7.78	83.12±6.35	80.02±5.88	81.84±4.25
WebKB	ACC	67.86±4.10	55.79±5.19	70.59±6.63	76.87±1.50	75.37±8.65	76.92±5.86	73.77±0.22	74.31±2.85	**77.51±0.66**
	NMI	37.63±3.49	12.14±5.56	40.79±4.41	43.07±3.00	48.90±5.84	48.06±3.59	40.22±4.71	45.87±5.21	**48.95±4.75**
	ARI	40.45±5.80	10.35±8.11	46.08±4.75	52.68±11.71	55.83±10.99	**57.23±7.84**	43.35±5.53	55.14±4.41	55.85±3.01
	Fscore	64.48±2.18	54.89±0.66	66.74±4.18	72.95±4.97	73.34±8.18	**74.80±5.82**	65.55±0.29	73.07±3.30	74.17±1.97
yale	ACC	39.79±2.65	43.64±2.69	41.45±2.97	49.91±4.89	48.33±4.74	54.42±5.97	50.94±4.60	55.15±8.14	**62.33±4.15**
	NMI	47.03±2.81	49.28±2.90	46.14±1.75	54.58±4.46	56.42±3.77	60.06±4.49	57.05±3.14	60.89±6.29	**66.20±3.07**
	ARI	19.48±2.95	23.26±1.92	18.67±2.43	30.03±5.16	31.76±4.38	36.48±5.51	31.79±3.76	37.42±8.48	**43.55±5.13**
	Fscore	25.33±2.43	28.60±1.78	24.21±2.23	34.84±4.72	36.59±3.87	40.78±6.47	36.59±3.33	41.68±7.83	**47.29±4.72**
3sources	ACC	43.22±4.47	42.01±6.40	49.82±7.33	46.63±5.63	50.06±7.08	47.49±9.09	44.88±5.99	49.26±7.33	**52.16±8.78**
	NMI	20.19±6.33	17.70±7.24	25.95±9.05	29.00±8.86	**33.41±7.38**	26.03±8.94	30.34±4.51	28.70±7.76	31.47±7.51
	ARI	6.24±8.46	6.20±8.14	15.08±11.80	14.22±11.25	20.56±12.41	13.60±15.17	13.17±6.98	18.20±14.10	**24.45±15.13**
	Fscore	36.46±1.06	36.73±3.89	41.85±7.55	40.25±6.49	44.07±8.10	42.04±7.49	38.17±2.15	43.28±7.97	**48.11±8.11**
Caltech101-7	ACC	53.48±2.12	61.27±6.67	54.79±5.70	**65.14±2.82**	59.58±4.12	61.09±1.12	56.84±1.50	58.46±0.92	63.24±9.28
	NMI	32.91±1.80	54.47±3.96	32.69±1.53	37.78±1.57	51.08±3.73	51.07±2.51	35.24±1.12	47.38±4.16	**55.20±5.12**
	ARI	31.76±3.27	50.19±7.47	39.67±4.45	53.75±3.09	51.08±0.67	51.41±0.28	46.99±0.40	48.15±0.53	**61.89±3.01**
	Fscore	50.96±3.66	63.46±6.40	56.13±4.26	68.74±2.51	68.04±0.25	68.20±0.18	64.74±0.29	65.53±0.25	**74.12±2.50**



(a) MSRCV1 (b) outdoor scene (c) ORL

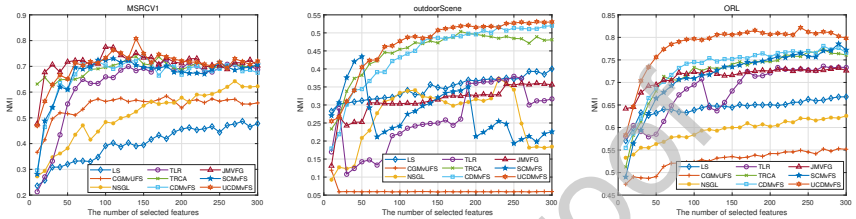
(d) BBCSport (e) handwritten (f) WebKB

(g) yale (h) 3sources (i) Caltech101-7

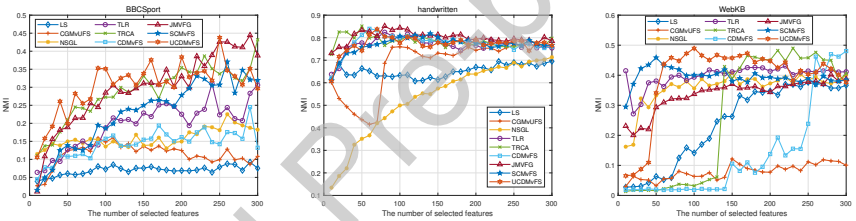
Fig 2: The ACC for different datasets with varying numbers of selected features across different algorithms.

Table 4: A comparison of different methods' runtime (in seconds) on nine real datasets.

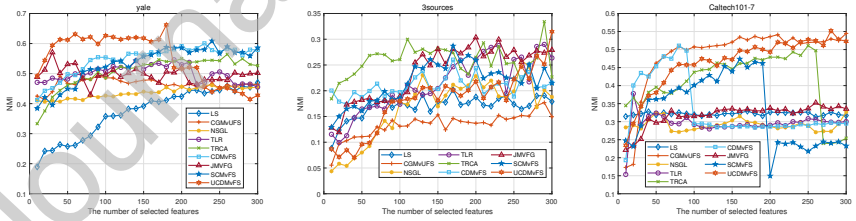
Method	3sources	BBCSport	Caltech101-7	handwritten	MSRCv1	ORL	outdoor scene	WebKB	yale
LS	0.0328	0.0216	0.1000	0.0348	0.0088	0.1164	0.0873	0.012	0.0489
NSGL	1092.4439	106.0148	217.0699	6.6706	12.1167	3063.9216	51.7113	15.6533	3094.4092
CGMvFS	0.6235	0.8771	4.3808	6.7079	0.2167	1.5597	8.4180	0.2365	0.6392
TLR	52.1377	20.9521	37.7320	37.9590	1.2849	168.6236	72.2404	2.4695	89.9407
TRCA-CGL	77.6280	39.3632	58.0721	68.2363	1.9069	300.9628	102.3549	3.0358	178.5176
CDMvFS	14.8899	11.6000	88.0649	232.8406	0.7885	38.9397	276.3153	1.4592	34.0289
JMVFG	22.6183	18.4808	22.9048	25.6867	0.6527	79.9890	53.7736	1.5575	57.0595
SCMvFS	24.8455	14.5361	67.4799	162.3849	5.1582	52.2985	255.0909	6.3539	41.4304
UCDMvFS	21.8836	13.5236	113.9488	238.1430	1.3251	44.2950	338.6632	1.7661	37.8672



(a) MSRCV1 (b) outdoor scene (c) ORL



(d) BBCSport (e) handwritten (f) WebKB



(g) yale (h) 3sources (i) Caltech101-7

Fig 3: The NMI for different datasets with varying numbers of selected features across different algorithms.

the best or second-best performance. The TRCA, CDMvFS, JMVFG, and SCMvFS methods also demonstrate the good performance. Figs. 2 and 3 illustrate the detailed performance variations across different feature selection ranges. Our analyses are as follows.

1. UCDMvFS demonstrates the strong performance on the two face datasets, outperforming the second-best algorithm by 5.71%, 3.58%, 6.21%, and 6.05% on ORL and by 7.18%, 5.31%, 6.13%, and 5.61% on yale across the four evaluation metrics, respectively. In terms of ACC, our method achieves improvements of 2.4%, 1.51%, 0.59%, and 2.1% on the MSRC-v1, Outdoor-Scene, WebKB, and 3source datasets, respectively.
2. By comparing the results, we found that UCDMvFS did not show good performance on the handwriting dataset, but TRCA-CGL, JMVFG and CDMvFS were able to show good results. The comparison methods that performed better than UCDMvFS all tried to find a consistent pseudo-label matrix, which means that for the handwriting dataset, a view-invariant label matrix may be a better choice. However, a view-invariant label matrix may ignore the unique information of different views. Therefore, we propose to preserve the uniqueness of the label matrix between multiple different views.
3. Among the compared algorithms, TRCA, CDMvFS, JMVFG, and SCMvFS can show the best or second-best performance on a few datasets, indicating that these methods are superior. However, compared with our method, tensor learning methods such as TLR and TRCA ignore the influence of divergent information between multiple views, and the latest graph learning methods such as CDMvFS, JMVFG, and SCMvFS ignore the higher-order relevance information between views and the cross-influence of information between different views. Overall, our method is effective.
4. In terms of computational efficiency, NSGL takes the most time. In addition, TLR, TRCA, CDMvFS, JMVFG, and SCMvFS also have large time costs on a few datasets. Although our method is effective, it introduces significant computational overhead due to the similarity matrix calculation or high-dimensional

space optimization, resulting in an overall time complexity of $O(n^3 + n^2V^2 + 2n^2V\log(n))$. This makes it difficult to apply to actual large-scale tasks.

5. By observing Figs. 2 and 3, we can find that on datasets with fewer features, such as webkb and handwritten, selecting about 50 highly representative features can achieve the best clustering effect and show strong dimensionality reduction capabilities. On datasets with more features, the best clustering effect among all the compared algorithms can be achieved within 300 features. Therefore, the method we proposed is effective.

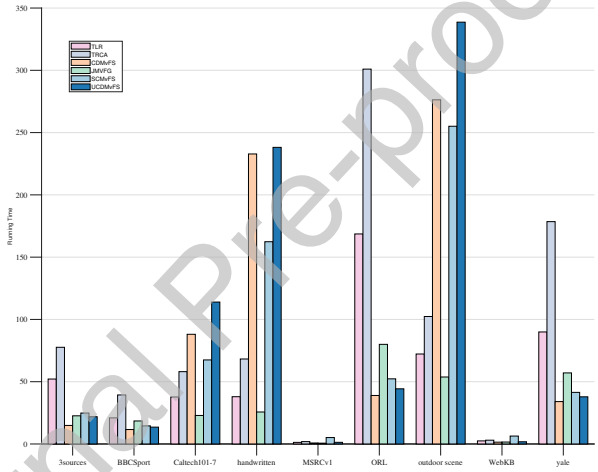
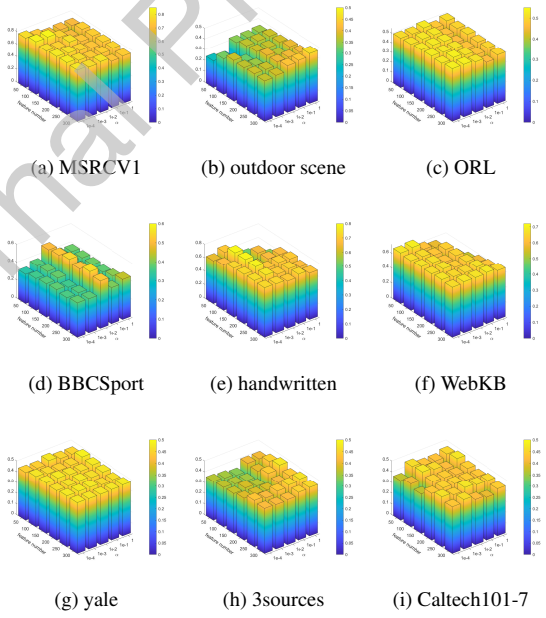


Fig 4: Relative running time of multi-view methods.

To further compare algorithm's performance, Table 4 presents the runtime of all methods across nine datasets. It is observed that our method requires more computational time on datasets where the number of samples exceeds the number of features, such as in the handwritten, Caltech101-7, and outdoor scene datasets. In contrast, it requires less time on datasets where the number of features exceeds the number of samples, such as ORL, 3sources, and yale. To illustrate the comparative results more clearly, we selected fives of the most representative recent methods along with our algorithm and plotted their runtime as bar charts, as shown in Fig. 4.

4.5. Ablation Study

In order to study whether tensor learning can obtain more comprehensive information, whether the $\ell_{2,1}$-norm of the feature selection matrix and the corresponding constraints are effective, we conducted ablation experiments and studies. First, we removed the tensor module in the UCDMvFS model and degenerated the model into UCDMvFS.woT. Secondly, we changed the $\ell_{2,1}$-norm in the UCDMvFS model to the F-norm and removed the corresponding constraints to form UCDMvFS.H.F-norm. The experimental results are shown in Table 5. From the data in the table, it can be seen that the more views the dataset has, the more obvious the effect of tensor learning is. On datasets with a small number of views, tensor learning can also improve certain performance. As for the effectiveness of the feature selection matrix constraints, it can be clearly seen from the table that the $\ell_{2,1}$-norm is better than the F-norm. Therefore, both tensor learning and the $\ell_{2,1}$-norm can effectively improve the performance of the algorithm.



(a) MSRCV1 (b) outdoor scene (c) ORL

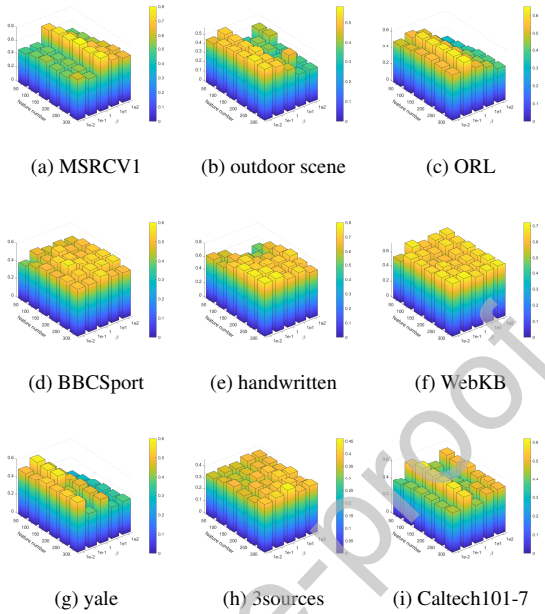
(d) BBCSport (e) handwritten (f) WebKB

(g) yale (h) 3sources (i) Caltech101-7

Fig 5: Effect of different values of α for clustering performance.

Table 5: Ablation studies on different datasets.

Datasets	Method	ACC	NMI	ARI	Fscore
MSRC_v1	UCDMvFS.woT	82.14±7.08	76.54±8.01	69.28±8.73	73.79±7.27
	UCDMvFS.H.F-norm	69.12±6.07	62.54±4.06	52.57±7.08	59.50±5.55
	UCDMvFS	**85.69±8.71**	**80.78±6.64**	**75.26±9.68**	**78.81±8.20**
outdoor_scene	UCDMvFS.woT	59.75±3.85	51.36±2.16	40.41±2.44	48.27±2.10
	UCDMvFS.H.F-norm	36.23±3.32	21.20±3.67	15.18±3.02	27.72±2.37
	UCDMvFS	**64.32±3.42**	**53.09±0.87**	**43.35±1.17**	**50.73±0.84**
ORL	UCDMvFS.woT	65.08±3.25	81.13±2.42	53.30±5.35	54.50±5.18
	UCDMvFS.H.F-norm	42.25±2.32	64.52±2.29	25.33±3.05	27.46±2.88
	UCDMvFS	**66.91±3.43**	**82.17±1.59**	**55.23±3.86**	**56.36±3.75**
BBCSport	UCDMvFS.woT	61.55±10.68	43.38±12.85	**34.27±13.86**	**54.43±8.44**
	UCDMvFS.H.F-norm	54.74±8.92	36.30±8.91	23.92±12.46	48.00±7.27
	UCDMvFS	**62.07±8.51**	**43.79±12.25**	33.56±14.14	54.14±10.98
handwritten	UCDMvFS.woT	84.04±6.26	82.79±2.89	78.20±5.06	80.47±4.47
	UCDMvFS.H.F-norm	76.51±5.90	73.24±2.79	65.92±5.48	69.43±4.87
	UCDMvFS	**87.46±5.84**	**83.77±3.58**	**79.78±4.79**	**81.84±4.25**
WebKB	UCDMvFS.woT	77.00±1.57	42.88±5.95	**55.95±3.21**	**74.29±2.01**
	UCDMvFS.H.F-norm	74.83±4.11	39.96±3.32	47.63±7.55	69.22±5.28
	UCDMvFS	**77.51±0.66**	**48.95±4.75**	55.85±3.01	74.17±1.97
yale	UCDMvFS.woT	62.06±3.72	65.31±3.87	41.56±5.61	45.54±5.14
	UCDMvFS.H.F-norm	38.97±2.54	45.93±2.58	18.50±2.84	24.46±2.43
	UCDMvFS	**62.33±4.15**	**66.20±3.07**	**43.55±5.13**	**47.29±4.72**
3sources	UCDMvFS.woT	47.04±8.21	29.07±6.97	15.37±13.64	41.43±8.32
	UCDMvFS.H.F-norm	51.98±7.24	31.88±5.03	20.70±9.06	44.71±7.83
	UCDMvFS	**52.16±8.78**	**31.47±7.51**	**24.45±15.13**	**48.11±8.11**
Caltech101-7	UCDMvFS.woT	60.03±3.67	48.70±4.33	49.64±6.44	66.45±5.94
	UCDMvFS.H.F-norm	54.56±1.51	33.88±1.14	33.12±1.25	52.60±1.53
	UCDMvFS	**63.24±9.28**	**55.20±5.12**	**61.89±3.01**	**74.12±2.50**



(a) MSRCV1 (b) outdoor scene (c) ORL

(d) BBCSport (e) handwritten (f) WebKB

(g) yale (h) 3sources (i) Caltech101-7

Fig 6: Effect of different values of β for clustering performance.

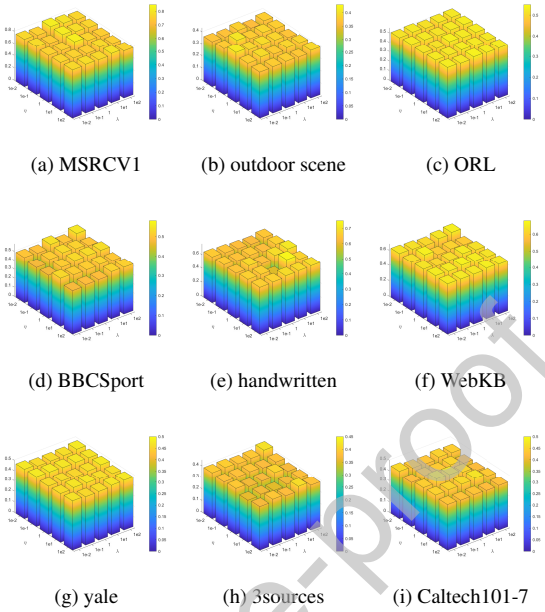
4.6. Parameter sensitivity

To further evaluate the performance of the proposed algorithm, we conducted parameter sensitivity experiments. For parameters α and β, we examined their impact on the model's performance by fixing the other three parameters at their median value of 1. Similarly, for parameters λ and η, we explored their influence by setting α and β to 1 while fixing the number of selected features at 150. The parameter sensitivity results are visualized using 3D bar charts, as shown in Figs. 5, 6, and 7.

We observed that variations in parameters λ and η have minimal impact on the model's performance. The model is more sensitive to parameter α, indicating that it should be adjusted based on the specific dataset. In contrast, parameter β should be set to a relatively small value to achieve the optimal model performance.

4.7. Convergence Study

This section further validates the model's convergence through experimental results. Fig. 8 illustrates the variation in the objective function value with the number



(a) MSRCV1 (b) outdoor scene (c) ORL

(d) BBCSport (e) handwritten (f) WebKB

(g) yale (h) 3sources (i) Caltech101-7

Fig 7: Effect of different values of λ and η for clustering performance.

of iterations for UCDMVFS across nine datasets. As shown in the figure, the objective function stabilizes within the first six iterations, indicating the rapid convergence. Although the convergence performance of the Caltech101-7 dataset increases slightly from three to five iterations, which may be due to reaching a local minimum rather than a global minimum, it eventually converges. Thus, the model's convergence is effectively confirmed.

5. Conclusion

In this paper, we employ a unified module to jointly measure multi-view consistency and diversity, thereby obtaining the pure graph for each view. These pure graphs are then fused to generate a consensus graph, with the optimal pure graphs learned through mutual constraint and promotion via self-weight learning and latent representation. Additionally, we propose using low-rank tensors to capture high-order graph structure information across multiple views. Finally, feature selection is performed

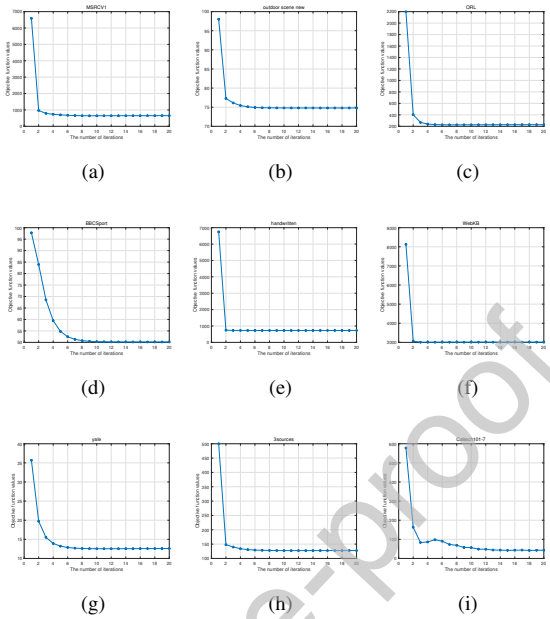


Fig 8: The convergence curves of the UCDMvFS algorithm on nine datasets.

alongside structural learning. These methods are seamlessly integrated within a single model framework, and the objective function is minimized using ADMM. This method has potential for data cleaning and data screening in medical imaging, computer vision, and media data applications. We found that unified measurement learning can effectively promote mutual learning between graphs, while tensor learning can explore high-order correlations between views and fully utilize the advantages of multiple views to improve model capabilities. The $\ell_{2,1}$-norm maintains the sparsity of the selected features and greatly reduces the interference of redundant information. In future work, we will focus on addressing the issue of parameter sensitivity, optimizing the inference speed, and reducing the memory footprint of the model for deployment on edge devices will be an important next step. Moreover, the pre-defined local structure similarity graph limits its ability to adaptively capture and describe local relationships in the data. Exploring how to dynamically update this graph will be a key direction for future research.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by National Natural Science Foundation of China(No.61906101).

It is also supported by the Ningbo Municipal Natural Science Foundation of China (No. 2023J115).

References

[1] G. Chao, K. Xu, X. Xie, Y. Chen, Global graph propagation with hierarchical information transfer for incomplete contrastive multi-view clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 39, pp. 15713–15721.

[2] H. Khalid, M. Hussain, M. A. Al Ghamdi, T. Khalid, K. Khalid, M. A. Khan, K. Fatima, K. Masood, S. H. Almotiri, M. S. Farooq, et al., A comparative systematic literature review on knee bone reports from mri, x-rays and ct scans using deep learning and machine learning methodologies, Diagnostics 10 (8) (2020) 518.

[3] L. Li, M. Cai, Drug target prediction by multi-view low rank embedding, IEEE/ACM transactions on computational biology and bioinformatics 16 (5) (2017) 1712–1721.

[4] X. Yu, M. Marinov, A study on recent developments and issues with obstacle detection systems for automated vehicles, Sustainability 12 (8) (2020) 3281.

[5] Z. Yang, Y. Tan, T. Yang, Large-scale multi-view clustering via matrix factorization of consensus graph, Pattern Recognition 155 (2024) 110716.

[6] Q. Ou, P. Zhang, S. Zhou, E. Zhu, One-step late fusion multi-view clustering with compressed subspace, in: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2024, pp. 7765–7769.

[7] Z. Cao, X. Xie, Multi-view unsupervised complementary feature selection with multi-order similarity learning, Knowledge-Based Systems 283 (2024) 111172.

[8] Z. Cao, X. Xie, Structure learning with consensus label information for multi-view unsupervised feature selection, Expert Systems with Applications 238 (2024) 121893.

[9] B. Jiang, C. Zhang, Z. Wang, X. Liang, P. Zhou, L. Du, Q. Zhang, W. Ding, Y. Liu, Scalable fuzzy clustering with collaborative structure learning and preservation, IEEE Transactions on Fuzzy Systems 33 (9) (2025) 3047–3060.

[10] Z. Lai, F. Chen, J. Wen, Multi-view robust regression for feature extraction, Pattern Recognition 149 (2024) 110219.

[11] C. Wang, J. Wang, Z. Gu, J.-M. Wei, J. Liu, Unsupervised feature selection by learning exponential weights, Pattern Recognition 148 (2024) 110183.

[12] Y. Wang, X. Li, J. Wang, A neurodynamic optimization approach to supervised feature selection via fractional programming, Neural Networks 136 (2021) 194–206.

[13] R. Sheikhpour, M. A. Sarram, S. Gharaghani, M. A. Z. Chahooki, A survey on semi-supervised feature selection methods, Pattern Recognition 64 (2017) 141–158.

[14] B. Jiang, J. Liu, Z. Wang, C. Zhang, J. Yang, Y. Wang, W. Sheng, W. Ding, Semi-supervised multi-view feature selection with adaptive similarity fusion and learning, Pattern Recognition 159 (2025) 111159.

[15] B. Jiang, X. Wu, X. Zhou, A. G. Cohn, Y. Liu, W. Sheng, H. Chen, Semi-supervised multi-view feature selection with adaptive graph learning, IEEE

Transactions on Neural Networks and Learning Systems 35 (3) (2024) 3615–3629.

[16] P. Huang, X. Yang, Unsupervised feature selection via adaptive graph and dependency score, Pattern Recognition 127 (2022) 108622.

[17] S. Liang, Q. Xu, P. Zhu, Q. Hu, C. Zhang, Unsupervised feature selection by manifold regularized self-representation, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017, pp. 2398–2402.

[18] C. Zhang, Y. Fang, X. Liang, H. Zhang, P. Zhou, X. Wu, J. Yang, B. Jiang, W. Sheng, Efficient multi-view unsupervised feature selection with adaptive structure learning and inference, in: Proceedings of the 33rd International Joint Conference on Artificial Intelligence, 2024, pp. 5443–5452.

[19] C. Tang, J. Chen, X. Liu, M. Li, P. Wang, M. Wang, P. Lu, Consensus learning guided multi-view unsupervised feature selection, Knowledge-Based Systems 160 (2018) 49–60.

[20] Y. Liu, K. Liu, C. Zhang, J. Wang, X. Wang, Unsupervised feature selection via diversity-induced self-representation, Neurocomputing 219 (2017) 350–363.

[21] C. Tang, X. Zheng, X. Liu, W. Zhang, J. Zhang, J. Xiong, L. Wang, Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection, IEEE Transactions on Knowledge and Data Engineering 34 (10) (2021) 4705–4716.

[22] G. Jiang, J. Peng, H. Wang, Z. Mi, X. Fu, Tensorial multi-view clustering via low-rank constrained high-order graph learning, IEEE Transactions on Circuits and Systems for Video Technology 32 (8) (2022) 5307–5318.

[23] Z. Wang, Q. Lin, Y. Ma, X. Ma, Local high-order graph learning for multi-view clustering, IEEE Transactions on Big Data (2024).

[24] Z. Li, C. Tang, X. Liu, X. Zheng, W. Zhang, E. Zhu, Consensus graph learning for multi-view clustering, IEEE Transactions on Multimedia 24 (2021) 2461–2472.

[25] X. Sang, J. Lu, H. Lu, Consensus graph learning for auto-weighted multi-view projection clustering, Information Sciences 609 (2022) 816–837.

[26] S. Huang, I. W. Tsang, Z. Xu, J. Lv, Measuring diversity in graph learning: A unified framework for structured multi-view clustering, IEEE Transactions on Knowledge and Data Engineering 34 (12) (2021) 5869–5883.

[27] Z. Chen, P. Lin, Z. Chen, D. Ye, S. Wang, Diversity embedding deep matrix factorization for multi-view clustering, Information Sciences 610 (2022) 114–125.

[28] Z. Li, C. Tang, J. Chen, C. Wan, W. Yan, X. Liu, Diversity and consistency learning guided spectral embedding for multi-view clustering, Neurocomputing 370 (2019) 128–139.

[29] S. Huang, Y. Liu, I. W. Tsang, Z. Xu, J. Lv, Multi-view subspace clustering by joint measuring of consistency and diversity, IEEE Transactions on Knowledge and Data Engineering 35 (8) (2022) 8270–8281.

[30] F. Zhang, H. Che, Separable consistency and diversity feature learning for multi-view clustering, IEEE Signal Processing Letters (2024).

[31] W. Hao, S. Pang, B. Yang, J. Xue, Tensor-based multi-view clustering with consistency exploration and diversity regularization, Knowledge-Based Systems 252 (2022) 109342.

[32] Y. Mi, Z. Ren, M. Mukherjee, Y. Huang, Q. Sun, L. Chen, Diversity and consistency embedding learning for multi-view subspace clustering, Applied Intelligence 51 (2021) 6771–6784.

[33] Z. Cao, X. Xie, Y. Li, Multi-view unsupervised feature selection with consensus partition and diverse graph, Information Sciences 661 (2024) 120178.

[34] Y. Huang, Z. Shen, Y. Cai, X. Yi, D. Wang, F. Lv, T. Li, C2imufs: Complementary and consensus learning-based incomplete multi-view unsupervised feature

selection, IEEE Transactions on Knowledge and Data Engineering 35 (10) (2023) 10681–10694.

[35] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, Advances in neural information processing systems 18 (2005).

[36] H. Bai, M. Huang, P. Zhong, Precise feature selection via non-convex regularized graph embedding and self-representation for unsupervised learning, Knowledge-Based Systems 296 (2024) 111900.

[37] C. Tang, X. Liu, M. Li, P. Wang, J. Chen, L. Wang, W. Li, Robust unsupervised feature selection via dual self-representation and manifold regularization, Knowledge-Based Systems 145 (2018) 109–120.

[38] X. Zhu, S. Zhang, R. Hu, Y. Zhu, et al., Local and global structure preservation for robust unsupervised spectral feature selection, IEEE Transactions on Knowledge and Data Engineering 30 (3) (2017) 517–529.

[39] J.-S. Wu, M.-X. Song, W. Min, J.-H. Lai, W.-S. Zheng, Joint adaptive manifold and embedding learning for unsupervised feature selection, Pattern Recognition 112 (2021) 107742.

[40] J. D. Carroll, J.-J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition, Psychometrika 35 (3) (1970) 283–319.

[41] L. R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika 31 (3) (1966) 279–311.

[42] M. E. Kilmer, K. Braman, N. Hao, R. C. Hoover, Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging, SIAM Journal on Matrix Analysis and Applications 34 (1) (2013) 148–172.

[43] Z. Zhang, G. Ely, S. Aeron, N. Hao, M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-svd, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 3842–3849.

[44] Y. Zhang, X. Wang, Z. Cai, Y. Zhou, S. Y. Philip, Tensor-based unsupervised multi-view feature selection for image recognition, in: 2021 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2021, pp. 1–6.

[45] L. Wang, C. Liang, W. Dong, W. Chen, Multi-view unsupervised feature selection via consensus guided low-rank tensor learning, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2022, pp. 575–580.

[46] H. Yuan, J. Li, Y. Liang, Y. Y. Tang, Multi-view unsupervised feature selection with tensor low-rank minimization, Neurocomputing 487 (2022) 75–85.

[47] C. Liang, L. Wang, L. Liu, H. Zhang, F. Guo, Multi-view unsupervised feature selection with tensor robust principal component analysis and consensus graph learning, Pattern Recognition 141 (2023) 109632.

[48] F. Nie, J. Li, X. Li, et al., Self-weighted multiview clustering with multiple graphs., in: IJCAI, 2017, pp. 2564–2570.

[49] F. Nie, X. Wang, M. Jordan, H. Huang, The constrained laplacian rank algorithm for graph-based clustering, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 30, 2016, pp. 1969–1976.

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

[51] W. Hu, D. Tao, W. Zhang, Y. Xie, Y. Yang, The twist tensor nuclear norm for video completion, IEEE transactions on neural networks and learning systems 28 (12) (2016) 2961–2973.

[52] P. Zhu, W. Zuo, L. Zhang, Q. Hu, S. C. Shiu, Unsupervised feature selection by regularized self-representation, Pattern Recognition 48 (2) (2015) 438–446.

[53] C. Tang, M. Bian, X. Liu, M. Li, H. Zhou, P. Wang, H. Yin, Unsupervised feature selection via latent representation learning and manifold regularization, Neural Networks 117 (2019) 163–178.

[54] X. Bai, L. Zhu, C. Liang, J. Li, X. Nie, X. Chang, Multi-view feature selection via nonnegative structured graph learning, Neurocomputing 387 (2020) 110–122.

[55] S.-G. Fang, D. Huang, C.-D. Wang, Y. Tang, Joint multi-view unsupervised feature selection and graph learning, IEEE Transactions on Emerging Topics in Computational Intelligence 8 (1) (2024) 16–31.