

# Structure-aware Hybrid-order Similarity Learning for Multi-view Unsupervised Feature Selection

Lin Xu, Ke Li, Dongjie Wang, Fengmao Lv, Tianrui Li, *Senior Member, IEEE*, and Yanyong Huang

**Abstract**—Multi-view unsupervised feature selection (MUFS) has recently emerged as an effective dimensionality reduction method for unlabeled multi-view data. However, most existing methods mainly use first-order similarity graphs to preserve local structure, often overlooking the global structure that can be captured by second-order similarity. In addition, a few MUFS methods leverage predefined second-order similarity graphs, making them vulnerable to noise and outliers and resulting in suboptimal feature selection performance. In this paper, we propose a novel MUFS method, termed Structure-aware Hybrid-order Similarity learning for multi-viEW unsupervised Feature Selection (SHINE-FS), to address the aforementioned problem. SHINE-FS first learns consensus anchors and the corresponding anchor graph to capture the cross-view relationships between the anchors and the samples. Based on the acquired cross-view consensus information, it generates low-dimensional representations of the samples, which facilitate the reconstruction of multi-view data by identifying discriminative features. Subsequently, it employs the anchor-sample relationships to learn a second-order similarity graph. Furthermore, by jointly learning first-order and second-order similarity graphs, SHINE-FS constructs a hybrid-order similarity graph that captures both local and global structures, thereby revealing the intrinsic data structure to enhance feature selection. Comprehensive experimental results on real multi-view datasets show that SHINE-FS outperforms the state-of-the-art methods.

**Index Terms**—Multi-view unsupervised feature selection, consensus anchor graph, hybrid-order similarity graph.

## I. INTRODUCTION

WITH the rapid advancement of information technology, multi-view data has become increasingly common in real-world applications, where each sample is described by heterogeneous features from different perspectives [1], [2]. However, the high dimensionality of multi-view data often results in the curse of dimensionality, significantly reducing the performance of downstream tasks [3], [4]. Moreover, labeling such data is difficult due to the high cost of manual annotation [5], [6]. To address these challenges, multi-view unsupervised feature selection (MUFS) effectively reduces the

Lin Xu, Ke Li, and Yanyong Huang are with the Joint Laboratory of Data Science and Business Intelligence, School of Statistics and Data Science, Southwestern University of Finance and Economics, Chengdu 611130, China, and also with the Big Data Laboratory on Financial Security and Behavior, SWUFE (Laboratory of Philosophy and Social Sciences, Ministry of Education), Chengdu 611130, China (e-mail: xulin00@foxmail.com; likec@swufe.edu.cn; huangyy@swufe.edu.cn), Yanyong Huang is the corresponding author;

Dongjie Wang is with the Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, USA (e-mail: wangdongjie@ku.edu);

Fengmao Lv and Tianrui Li are with the School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China (e-mail: fengmaolv@126.com; trli@swjtu.edu.cn).

dimensionality of unlabeled multi-view data by selecting a compact subset of informative features without relying on labels [7], [8].

Existing MUFS methods can be broadly divided into two categories. The first concatenates features from all views and then applies single-view feature selection methods, such as LPSCore [9], HSL [10], and GAWFS [11]. In contrast, the second category selects discriminative features directly from multi-view data by leveraging inter-view correlations. Representative methods include JMVFG [12], GCDUFS [13], and UKMFS [14]. Leveraging similarity graphs to preserve local structure has been shown to significantly improve MUFS performance, leading to the development of numerous graph-based approaches. JMVFG [12] employs orthogonal decomposition to obtain both view-specific basis matrices and a shared clustering indicator matrix, while simultaneously learning a consensus similarity graph to guide feature selection. GCDUFS [13] learns a consensus graph to capture cross-view information and derives diverse representations from different views to exploit their complementary information, both of which help identify discriminative features. UKMFS [14] unifies multi-view representations via nonlinear kernel mapping, learns a consistent graph under low-rank constraints, and utilizes binary hashing codes to facilitate feature selection.

Although existing graph-based methods have achieved promising performance, they typically rely on first-order similarity graphs to preserve local data structure [15], while often neglecting global structures captured by second-order similarity graphs [16]. For example, in document clustering, a first-order similarity graph connects only documents that are directly similar in content, thus making them immediate neighbors. Extending beyond direct pairwise similarities, a second-order similarity graph captures topic-level groupings by linking documents that share many common neighbors—even if they are not directly similar—since these documents are connected to the same set of other documents. This enables the second-order graph to capture more global relationships within the corpus [17]. A few studies have investigated unsupervised feature selection for second-order graphs. HSL [10] exploits first-order and higher-order similarities in the low-dimensional embedding space to capture structural information, and learns an optimal graph Laplacian for selecting discriminative features. CFSMO [18] projects multi-view data into a shared latent space to capture complementary information, and integrates multi-order structural information across multiple views to guide feature selection. These approaches typically predefine second-order graphs before performing feature selection. However, their vulnerability to noise and

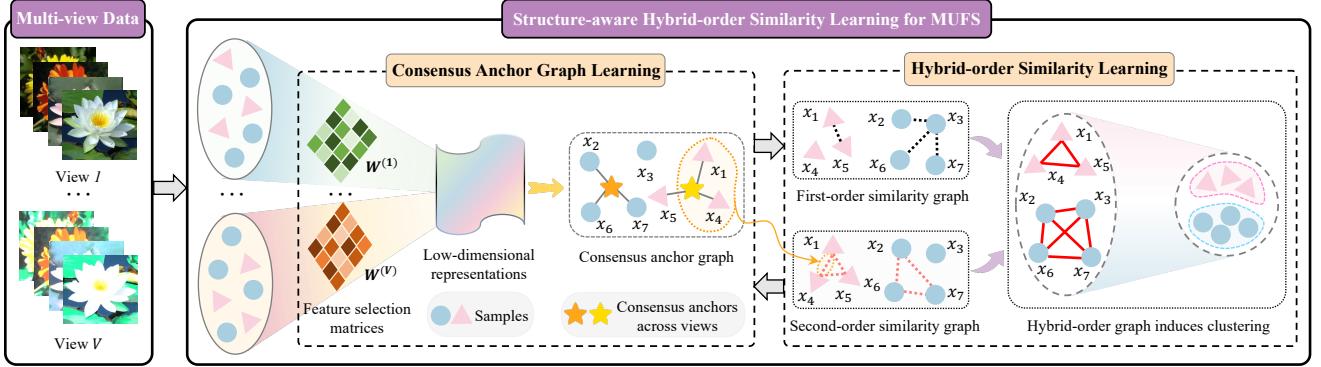


Fig. 1. The framework of the proposed SHINE-FS.

outliers often compromises feature selection performance.

To address the aforementioned issues, we propose a novel MUFS method called Structure-aware Hybrid-order Similarity learning for multi-viEW unsupervised Feature Selection (SHINE-FS). Specifically, we first learn a set of consensus anchors and the corresponding anchor graph to capture cross-view relationships between anchors and samples. Next, we derive low-dimensional representations of the samples from the learned consensus information, enabling the reconstruction of multi-view data and the identification of discriminative features. Subsequently, we learn a second-order similarity graph based on anchor-sample relationships within the consensus anchor graph, where samples strongly connected to the same anchor are considered to share a second-order relationship. Furthermore, by jointly learning first-order and second-order similarity graphs to construct a hybrid-order graph, SHINE-FS effectively captures both local and global structural information, thereby better preserving the intrinsic data structure and guiding feature selection. Finally, an iterative optimization algorithm is developed for the proposed model, and extensive experiments on multiple real-world multi-view datasets are conducted to demonstrate the effectiveness of SHINE-FS over state-of-the-art methods. The framework of SHINE-FS is illustrated in Fig. 1.

The main contributions of this work can be summarized as follows: (i) Instead of limiting MUFS to the use of a first-order similarity graph or simply combining it with a predefined second-order graph, we propose a hybrid-order similarity graph adaptive learning approach that incorporates consensus anchor graph-induced regularization, which fully leverages both local and global structural information from graphs of different orders to enhance feature selection. (ii) SHINE-FS seamlessly integrates multi-view feature selection, consensus anchor graph learning, and hybrid-order similarity learning into a unified framework, allowing these components to mutually reinforce each other and ultimately improve feature selection performance. (iii) We develop an effective and fast-converging iterative optimization algorithm to implement the proposed method. Extensive experimental results on real-world multi-view datasets demonstrate that SHINE-FS outperforms state-of-the-art approaches.

The remainder of this paper is organized as follows. Section II reviews related work on MUFS. In Section III, the proposed

SHINE-FS method is presented, together with the iterative optimization algorithm developed to address the formulated optimization problem, as well as analyses of the algorithm's time complexity and convergence. Extensive experimental results on real multi-view datasets are reported in Section IV. Section V concludes the paper.

## II. RELATED WORK

In this section, we briefly introduce some representative MUFS methods, which can be broadly categorized into single-view and multi-view unsupervised feature selection approaches.

To select informative features from multi-view data, the first category of MUFS approaches concatenates features from all views and subsequently conducts feature selection using conventional single-view methods, typically including LPSCore [9] and GAWFS [11]. LPSCore identifies informative features by using the Laplacian score, which evaluates how well each feature preserves the local structure of the data [9]. GAWFS combines non-negative matrix factorization with adaptive graph learning to discover discriminative features through a learned feature weighting matrix [11]. Jiang et al. [19] developed a graph-based unsupervised feature selection approach that constructs a block-diagonal graph within a low-dimensional subspace, thereby effectively capturing the intrinsic data structure to inform the feature selection process. The single-view-based methods mentioned above address multi-view data by simply merging features from different views, but they overlook correlations between views, which can result in suboptimal feature selection.

In contrast, the second category of MUFS methods directly selects discriminative features from multi-view data by leveraging inter-view correlations to enhance performance. Zhou et al. [20] introduced a structural regularization-based MUFS method that preserves data locality by jointly learning both view-specific and consensus similarity graphs throughout the feature selection process. JMVFG [12] maintains cross-view consistency by decomposing the projected data matrix into view-specific basis matrices and a clustering indicator matrix, while simultaneously learning a consensus graph that preserves the local data structure to facilitate feature selection. Yuan et al. [21] promoted cross-view consistency through tensor low-rank regularization and utilized intra-view structural

information from view-specific graphs for feature selection. GCDUFS [13] captures consensus information across views via graph regularization and explores inter-view diversity using kernel-based dependency measures, integrating both aspects into the feature selection process. Zhang et al. [22] developed an efficient MUFS method that collaboratively fuses membership matrices and bipartite graphs to capture both cluster structure and local data structure for discriminative feature selection. UKMFS [14] unifies multi-view data representations through kernel mapping, learns a consistent graph across views using low-rank constraints, and employs binary hashing codes to guide feature selection. Cao et al. [23] utilized view-specific similarity graphs to explore a consensus clustering structure, aiding feature selection by leveraging both intra-view and inter-view information. CCSFS [24] identifies discriminative features by learning a consensus cluster indicator matrix, which is obtained by fusing partition matrices from multiple views.

Although these methods achieve promising results in feature selection, they typically preserve the local data structure by utilizing first-order similarity graphs, while often overlooking the global structures captured by second-order graphs. Only a few studies have explored the use of second-order graphs for feature selection. HSL [10] projects high-dimensional data into a low-dimensional space, preserving both first-order and higher-order similarity structures to facilitate the identification of discriminative features. CFSMO [18] selects features by integrating neighbor information from multi-order similarity graphs with consistency information from a shared latent representation. The two methods mentioned above typically predefine second-order graphs, making them vulnerable to noise and outliers and ultimately diminishing the effectiveness of feature selection.

### III. THE PROPOSED METHOD

In this section, we begin by summarizing the notations and definitions used throughout the paper. We then introduce our proposed method SHINE-FS, and develop an iterative algorithm to solve the associated optimization problem. Finally, we analyze the time complexity and convergence of the algorithm.

#### A. Notations and Definitions

Throughout this paper, matrices are represented by boldface uppercase letters. For a matrix  $\mathbf{M} \in \mathbb{R}^{p \times q}$ ,  $\mathbf{M}_{\cdot i}$  and  $\mathbf{M}_{i \cdot}$  denote its  $i$ -th row and  $j$ -th column, respectively, and  $M_{ij}$  represents the  $(i, j)$ -th entry. The transpose of  $\mathbf{M}$  is denoted by  $\mathbf{M}^T$ , and its trace is  $\text{Tr}(\mathbf{M})$ . The Frobenius norm of  $\mathbf{M}$  is defined as  $\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^p \sum_{j=1}^q M_{ij}^2}$ , and the  $\ell_{2,1}$ -norm is given by  $\|\mathbf{M}\|_{2,1} = \sum_{i=1}^p \sqrt{\sum_{j=1}^q M_{ij}^2}$ . Let  $\|\mathbf{M}\|_0$  denote the  $\ell_0$ -norm of  $\mathbf{M}$ , which counts the number of non-zero elements in  $\mathbf{M}$ .  $\mathbf{I}$  denotes the identity matrix.  $\mathbf{1}_n = [1, \dots, 1]^T$  is an  $n$ -dimensional column vector of ones. It is denoted by  $\mathbf{1}$  when the size is clear from the context. Let  $\mathcal{X} = \{\mathbf{X}^{(v)}\}_{v=1}^l$  represent a multi-view dataset with  $l$  views, where  $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$  denotes the data matrix for the  $v$ -th view, consisting of  $n$  instances and  $d_v$  features. Our goal is to

select the  $h$  most discriminative features from the multi-view dataset  $\mathcal{X}$ .

#### B. Formulation of SHINE-FS

Previous studies have shown that consensus information among multiple views is essential for preserving their shared structure, which in turn facilitates feature selection [13], [25], [26]. Existing anchor-based MUFS methods typically select representative samples as anchors to construct view-specific anchor graphs, which are then fused to reveal the consensus structure [15], [22]. However, noise in the original feature space can prevent these predefined anchors from accurately capturing the underlying data structure. Additionally, these methods do not leverage anchor-sample relationships across views to enable second-order graph learning, which is vital for uncovering the global structure of the data [16]. To address these limitations, we propose an adaptive anchor-induced second-order graph learning approach to guide multi-view feature selection, as described below:

$$\begin{aligned} & \min_{\alpha^{(v)}, \mathbf{W}^{(v)}, \mathbf{C}, \mathbf{A}, \mathbf{S}} \sum_{v=1}^l (\alpha^{(v)})^2 [\|\mathbf{X}^{(v)} - \mathbf{W}^{(v)} \mathbf{C} \mathbf{A}\|_F^2 + \gamma \|\mathbf{W}^{(v)}\|_{2,1}] \\ & \quad + \sum_{i,j=1}^n \|\mathbf{A}_{\cdot i} - \mathbf{A}_{\cdot j}\|_2^2 S_{ij} + \beta \|\mathbf{A}\|_F^2 + \lambda_S \|\mathbf{S}\|_F^2 \\ & \text{s.t. } 0 \leq \alpha^{(v)} \leq 1, \sum_{v=1}^l \alpha^{(v)} = 1, \mathbf{W}^{(v)T} \mathbf{W}^{(v)} = \mathbf{I}, \mathbf{C}^T \mathbf{C} = \mathbf{I}, \\ & \quad \mathbf{A}^T \mathbf{1} = \mathbf{1}, A_{ij} \geq 0, \|\mathbf{S}_{\cdot i}\|_0 = k, \mathbf{S}_{\cdot i} \cdot \mathbf{1} = 1, S_{ij} \geq 0, \end{aligned} \quad (1)$$

where  $\alpha^{(v)}$  denotes the weight of the  $v$ -th view,  $\mathbf{C} \in \mathbb{R}^{c \times m}$  represents a set of  $m$  consensus anchors, each being a  $c$ -dimensional vector derived from the unified latent space,  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is the consensus anchor graph, and  $\mathbf{S} \in \mathbb{R}^{n \times n}$  represents the anchor-induced second-order similarity graph. By leveraging the direct anchor-sample relationships across views in  $\mathbf{A}$ ,  $\mathbf{S}$  captures indirect second-order relationships by linking samples that share common anchors, in accordance with the second-order graph principle whereby a neighbor's neighbor is also considered a neighbor [27]. Additionally,  $\mathbf{C} \mathbf{A}$  yields low-dimensional representations of the samples, wherein each sample is characterized by its associations to the consensus anchors. The matrix  $\mathbf{W}^{(v)} \in \mathbb{R}^{d_v \times c}$  facilitates the reconstruction of the original data in the  $v$ -th view from its corresponding low-dimensional representations, thereby quantifying the contribution of each feature to the reconstruction process. Furthermore, applying the  $\ell_{2,1}$ -norm regularization on  $\mathbf{W}^{(v)}$  promotes row-wise sparsity, helping to suppress redundant or irrelevant features while retaining the important ones.

In Eq. (1), rather than relying on predefined anchors, we adaptively learn a set of consensus anchors  $\mathbf{C}$  and the corresponding anchor graph  $\mathbf{A}$ , which together capture anchor-sample relationships across multiple views. From the learned consensus information, we derive low-dimensional representations of the samples. These representations are then used to reconstruct the original multi-view data by identifying

discriminative features. In addition, samples that share similar anchor neighbors can preserve their second-order relationships, thereby enabling effective characterization of the global data structure.

To leverage both local and global structural information for feature selection, we propose an adaptive hybrid-order similarity learning method that combines first-order and second-order similarity graphs, as detailed below:

$$\begin{aligned} & \min_{\mathbf{G}, \mathbf{F}, \alpha^{(v)}} \sum_{v=1}^l (\alpha^{(v)})^2 \operatorname{Tr}(\mathbf{X}^{(v)} \mathbf{L}_G \mathbf{X}^{(v)T}) + \lambda_G \|\mathbf{G}\|_F^2 \\ & + \frac{1}{2} \sum_{i,j=1}^n \|\mathbf{F}_{i \cdot} - \mathbf{F}_{j \cdot}\|_2^2 (\mathbf{G} + \eta \mathbf{S})_{ij} \\ \text{s.t. } & 0 \leq \alpha^{(v)} \leq 1, \sum_{v=1}^l \alpha^{(v)} = 1, \|\mathbf{G}_{i \cdot}\|_0 = k, \mathbf{G}_{i \cdot} \mathbf{1} = 1, \\ & G_{ij} \geq 0, \mathbf{F}^T \mathbf{F} = \mathbf{I}, \end{aligned} \quad (2)$$

where  $\mathbf{G}$  denotes the first-order similarity graph,  $\mathbf{L}_G$  is the corresponding Laplacian matrix,  $\mathbf{G} + \eta \mathbf{S}$  represents the hybrid-order similarity graph, and  $\mathbf{F}$  corresponds to the cluster indicator matrix. The Frobenius norm term  $\|\mathbf{G}\|_F^2$  is included to prevent trivial solutions, with  $\lambda_G$  serving as its regularization parameter. Additionally,  $\eta$  acts as a trade-off parameter, balancing the contributions of first-order and second-order similarities. In Eq. (2), the first-order similarity graph  $\lambda_G$  is adaptively learned to capture direct relationships among samples, thereby preserving the local structure of the data. Meanwhile, the anchor-induced second-order similarity graph  $\mathbf{S}$  utilizes shared anchor neighborhood information from the consensus anchor graph to identify second-order indirect correlations between samples, capturing the global data structure. These two graphs are then integrated into a hybrid-order graph, which jointly leverages both local and global structural information to better characterize the intrinsic structure of the data. The resulting hybrid-order graph further guides clustering, ensuring that highly similar samples are assigned similar pseudo-labels.

By combining Eqs. (1) and (2), the overall objective function of SHINE-FS is summarized as follows:

$$\begin{aligned} & \min_{\Omega} \sum_{v=1}^l (\alpha^{(v)})^2 [\|\mathbf{X}^{(v)} - \mathbf{W}^{(v)} \mathbf{C} \mathbf{A}\|_F^2 + \operatorname{Tr}(\mathbf{X}^{(v)} \mathbf{L}_G \mathbf{X}^{(v)T}) \\ & + \gamma \|\mathbf{W}^{(v)}\|_{2,1} + \sum_{i,j=1}^n \|\mathbf{A}_{i \cdot} - \mathbf{A}_{j \cdot}\|_2^2 S_{ij} + \beta \|\mathbf{A}\|_F^2 + \lambda_S \|\mathbf{S}\|_F^2] \\ & + \frac{1}{2} \sum_{i,j=1}^n \|\mathbf{F}_{i \cdot} - \mathbf{F}_{j \cdot}\|_2^2 (\mathbf{G} + \eta \mathbf{S})_{ij} + \lambda_G \|\mathbf{G}\|_F^2 \\ \text{s.t. } & 0 \leq \alpha^{(v)} \leq 1, \sum_{v=1}^l \alpha^{(v)} = 1, \mathbf{A}^T \mathbf{1} = \mathbf{1}, A_{ij} \geq 0, \\ & \|\mathbf{G}_{i \cdot}\|_0 = k, \mathbf{G}_{i \cdot} \mathbf{1} = 1, G_{ij} \geq 0, \mathbf{W}^{(v)T} \mathbf{W}^{(v)} = \mathbf{I}, \\ & \|\mathbf{S}_{i \cdot}\|_0 = k, \mathbf{S}_{i \cdot} \mathbf{1} = 1, S_{ij} \geq 0, \mathbf{F}^T \mathbf{F} = \mathbf{I}, \mathbf{C}^T \mathbf{C} = \mathbf{I}, \end{aligned} \quad (3)$$

where  $\Omega = \{\mathbf{C}, \mathbf{A}, \mathbf{G}, \mathbf{S}, \mathbf{F}, \{\mathbf{W}^{(v)}, \alpha^{(v)}\}_{v=1}^l\}$ . Here,  $\gamma$ ,  $\beta$ ,

and  $\eta$  are hyperparameters, whereas  $\lambda_S$  and  $\lambda_G$  are automatically determined during optimization of  $\mathbf{S}$  and  $\mathbf{G}$ , respectively.

Eq. (3) integrates multi-view feature selection, consensus anchor graph learning, and hybrid-order similarity learning into a unified framework. This unified framework leverages both local and global structural information to accurately characterize the intrinsic structure, thereby facilitating feature selection. Moreover, selecting discriminative features enhances the representativeness of the consensus anchors and the quality of the corresponding anchor graph, which in turn promotes the adaptive learning of hybrid-order similarity. These components mutually reinforce each other, thereby improving feature selection.

### C. Optimization and Algorithm

As the objective function in Eq. (3) is not jointly convex with respect to all variables, we propose an iterative optimization algorithm that alternately updates each variable while keeping the others fixed.

*1) Updating  $\mathbf{W}^{(v)}$  by Fixing Other Variables:* When the other variables are fixed, the optimization subproblem w.r.t.  $\mathbf{W}^{(v)}$  can be rewritten in the following equivalent trace form:

$$\begin{aligned} & \min_{\mathbf{W}^{(v)}} \operatorname{Tr}(\gamma \mathbf{W}^{(v)T} \mathbf{D}^{(v)} \mathbf{W}^{(v)} - 2 \mathbf{W}^{(v)T} \mathbf{X}^{(v)} \mathbf{A}^T \mathbf{C}^T) \\ & \text{s.t. } \mathbf{W}^{(v)T} \mathbf{W}^{(v)} = \mathbf{I}, \end{aligned} \quad (4)$$

where  $\mathbf{D}^{(v)}$  denotes a diagonal matrix with the  $i$ -th diagonal element defined as  $\mathbf{D}_{ii}^{(v)} = \frac{1}{2\sqrt{\|\mathbf{W}_{i \cdot}^{(v)}\|_2^2 + \varepsilon}}$  ( $\varepsilon$  is a small constant added to avoid division by zero). Eq. (4) can be efficiently solved using the generalized power iteration (GPI) method [28].

*2) Updating  $\mathbf{A}$  by Fixing Other Variables:* With other variables fixed, the subproblem of solving  $\mathbf{A}$  reduces to the following equivalent trace form:

$$\min_{\mathbf{A}^T \mathbf{1} = \mathbf{1}, A_{ij} \geq 0} \operatorname{Tr}(\mathbf{P} \mathbf{A}^T \mathbf{A} - \mathbf{Q} \mathbf{A}), \quad (5)$$

where  $\mathbf{P} = [\sum_{v=1}^l (\alpha^{(v)})^2 + \beta] \mathbf{I} + 2 \mathbf{L}_S$ , and  $\mathbf{Q} = \sum_{v=1}^l 2(\alpha^{(v)})^2 \mathbf{X}^{(v)T} \mathbf{W}^{(v)} \mathbf{C}$ . According to [22], the constraints on  $\mathbf{A}$  are first relaxed to obtain a latent solution  $\hat{\mathbf{A}}$  by setting the derivative of Eq. (5) to zero. Then, the latent solution is projected onto the constrained space as follows:

$$\min_{\mathbf{A}^T \mathbf{1} = \mathbf{1}, A_{ij} \geq 0} \|\mathbf{A} - \hat{\mathbf{A}}\|_F^2. \quad (6)$$

According to [29], the solution to Eq. (6) is given by:

$$\mathbf{A} = (\hat{\mathbf{A}} - \hat{\mathbf{A}} \mathbf{1}_n \mathbf{1}_n^T + \mathbf{1}_m \mathbf{1}_m^T, 0)_+. \quad (7)$$

*3) Updating  $\mathbf{C}$  by Fixing Other Variables:* By keeping the other variables fixed, solving  $\mathbf{C}$  is equivalent to the following problem:

$$\min_{\mathbf{C}^T \mathbf{C} = \mathbf{I}} \sum_{v=1}^l (\alpha^{(v)})^2 \|\mathbf{X}^{(v)} - \mathbf{W}^{(v)} \mathbf{C} \mathbf{A}\|_F^2. \quad (8)$$

Then, Eq. (8) can be transformed into an equivalent trace form:

$$\max_{\mathbf{C}^T \mathbf{C} = \mathbf{I}} \operatorname{Tr}(\mathbf{C}^T \mathbf{E}), \quad (9)$$

where  $\mathbf{E} = \sum_{v=1}^l 2(\alpha^{(v)})^2 \mathbf{W}^{(v)T} \mathbf{X}^{(v)} \mathbf{A}^T$ . The optimal solution to Eq. (9) can be obtained via the singular value decomposition (SVD) of  $\mathbf{E}$  as the product of its left and right singular vectors [30].

4) *Updating  $\mathbf{S}$  by Fixing Other Variables:* Since the rows of  $\mathbf{S}$  are independent,  $\mathbf{S}_i$  can be updated by solving the following equivalent problem after fixing the other variables:

$$\begin{aligned} & \min_{\mathbf{S}_i} \frac{1}{2} \|\mathbf{S}_i + \frac{\mathbf{U}_i}{2\lambda_S}\|_2^2 \\ & \text{s.t. } \|\mathbf{S}_i\|_0 = k, \mathbf{S}_i \cdot \mathbf{1} = 1, S_{ij} \geq 0, \end{aligned} \quad (10)$$

where  $\mathbf{U}_i$  denotes a vector, with its  $j$ -th entry  $U_{ij} = \sum_{r=1}^m (A_{ri} - A_{rj})^2 + \frac{\eta}{2} \|\mathbf{F}_i - \mathbf{F}_j\|_2^2$ . Following [31], the solution for  $\mathbf{S}$  is given by:

$$S_{ij} = \begin{cases} \frac{U_{i,k+1} - U_{ij}}{kU_{i,k+1} - \sum_{p=1}^k U_{ip}}, & j \leq k; \\ 0, & j > k. \end{cases} \quad (11)$$

In addition,  $\lambda_S$  is adaptively determined as  $(kU_{i,k+1} - \sum_{p=1}^k U_{ip})/2$  to ensure that  $\mathbf{S}_i$  contains exactly  $k$  nonzero elements.

5) *Updating  $\mathbf{G}$  by Fixing Other Variables:* Following a similar procedure as in Eq. (10), each row of  $\mathbf{G}$  is updated by solving the following equivalent optimization problem with the other variables fixed:

$$\begin{aligned} & \min_{\mathbf{G}_i} \frac{1}{2} \|\mathbf{G}_i + \frac{\mathbf{B}_i}{2\lambda_G}\|_2^2 \\ & \text{s.t. } \|\mathbf{G}_i\|_0 = k, \mathbf{G}_i \cdot \mathbf{1} = 1, G_{ij} \geq 0, \end{aligned} \quad (12)$$

where  $\mathbf{B}_i$  is a vector with the  $j$ -th element  $B_{ij} = \frac{1}{2} [\sum_{v=1}^l (\alpha^{(v)})^2 \|\mathbf{X}_i^{(v)} - \mathbf{X}_j^{(v)}\|_2^2 + \|\mathbf{F}_i - \mathbf{F}_j\|_2^2]$ . Similar to solving Eq. (10), the optimal solution for  $\mathbf{G}$  can be obtained as follows:

$$G_{ij} = \begin{cases} \frac{B_{i,k+1} - B_{ij}}{kB_{i,k+1} - \sum_{p=1}^k B_{ip}}, & j \leq k; \\ 0, & j > k. \end{cases} \quad (13)$$

Meanwhile,  $\lambda_G$  is set to  $(kB_{i,k+1} - \sum_{p=1}^k B_{ip})/2$  to guarantee that  $\mathbf{G}_i$  contains  $k$  nonzero elements.

6) *Updating  $\mathbf{F}$  by Fixing Other Variables:* With all other variables fixed, the optimization subproblem for  $\mathbf{F}$  can be formulated as follows:

$$\min_{\mathbf{F}^T \mathbf{F} = \mathbf{I}} \text{Tr}(\mathbf{F}^T \mathbf{L}^* \mathbf{F}), \quad (14)$$

where  $\mathbf{L}^*$  denotes the Laplacian matrix associated with the hybrid-order similarity graph. The solution to  $\mathbf{F}$  of Eq. (14) is given by the eigenvectors corresponding to the smallest  $c$  eigenvalues of  $\mathbf{L}^*$  [31].

7) *Updating  $\alpha^{(v)}$  by Fixing Other Variables:* When fixing other variables, we update  $\alpha^{(v)}$  by solving the following problem:

$$\begin{aligned} & \min_{\alpha^{(v)}} \sum_{v=1}^l (\alpha^{(v)})^2 z^{(v)} \\ & \text{s.t. } 0 \leq \alpha^{(v)} \leq 1, \sum_{v=1}^l \alpha^{(v)} = 1, \end{aligned} \quad (15)$$

---

**Algorithm 1** Iterative algorithm of SHINE-FS

---

**Input:** Multi-view data  $\mathcal{X} = \{\mathbf{X}^{(v)}\}_{v=1}^l$ , parameters  $\gamma$ ,  $\beta$ , and  $\eta$ .

- 1: Initialize:  $\{\mathbf{D}^{(v)}, \alpha^{(v)}\}_{v=1}^l$ ,  $\mathbf{A}$ ,  $\mathbf{C}$ ,  $\mathbf{S}$ , and  $\mathbf{G}$ .
- 2: **while** not convergent **do**
- 3:   Update  $\{\mathbf{W}^{(v)}\}_{v=1}^l$  by solving Eq. (4).
- 4:   Update  $\mathbf{A}$  according to Eq. (7).
- 5:   Update  $\mathbf{C}$  by solving Eq. (9).
- 6:   Update  $\mathbf{S}$  according to Eq. (11).
- 7:   Update  $\mathbf{G}$  according to Eq. (13).
- 8:   Update  $\mathbf{F}$  by solving Eq. (14).
- 9:   Update  $\{\alpha^{(v)}\}_{v=1}^l$  according to Eq. (16).
- 10: **end while**

**Output:** Sort the  $\ell_2$ -norms of the rows in  $\{\mathbf{W}^{(v)}\}_{v=1}^l$  in descending order, and select the top  $h$  features from  $\mathcal{X}$ .

---

where  $z^{(v)} = \|\mathbf{X}^{(v)} - \mathbf{W}^{(v)} \mathbf{C} \mathbf{A}\|_F^2 + \gamma \|\mathbf{W}^{(v)}\|_{2,1} + \text{Tr}(\mathbf{X}^{(v)} \mathbf{L}_G \mathbf{X}^{(v)T})$ . According to [5], we can obtain the optimal solution of  $\alpha^{(v)}$  as follows:

$$\alpha^{(v)} = \frac{(z^{(v)})^{-1}}{\sum_{v=1}^l (z^{(v)})^{-1}}. \quad (16)$$

The optimization procedure of SHINE-FS is summarized in Algorithm 1. In this algorithm,  $\mathbf{G}$  is initialized as a  $k$ -nearest neighbor graph constructed from the concatenated data of all views [32].  $\mathbf{A}$ ,  $\mathbf{C}$ , and  $\mathbf{S}$  are randomly initialized subject to their respective constraints. Moreover,  $\mathbf{D}^{(v)}$  and  $\alpha^{(v)}$  are initialized as the identity matrix and  $\frac{1}{l}$ , respectively.

#### D. Time Complexity and Convergence Analysis

Algorithm 1 involves the alternating update of seven variables. In each iteration,  $\mathbf{W}^{(v)}$  is updated using the GPI method with a computational complexity of  $\mathcal{O}(d_v^2 c)$ . Updating  $\mathbf{A}$  costs  $\mathcal{O}(n^2)$ . Updating  $\mathbf{C}$  takes  $\mathcal{O}(d_v cn + cnm + mc^2)$  complexity, where  $m \ll d_v$  and  $c \ll n$ . The computational complexity for updating  $\mathbf{F}$  is  $\mathcal{O}(n^2 c)$ . Updating  $\mathbf{S}$ ,  $\mathbf{G}$ , and  $\alpha^{(v)}$  involves only element-wise operations, so the computational cost is negligible. In summary, the time complexity per iteration of Algorithm 1 is  $\mathcal{O}(n^2 c + \sum_{v=1}^l d_v c(d_v + n))$ .

Since the objective function in Eq. (3) is non-convex with respect to all variables, we optimize it using an alternative iterative optimization algorithm, as described in Algorithm 1. Here, we provide a brief convergence analysis for each step. For step 3 of Algorithm 1, the convergence of the GPI method used to update  $\mathbf{W}^{(v)}$  is theoretically supported by [28]. In steps 4, 6, 7, and 9, the closed-form solutions to Eqs. (7), (11), (13), and (16) ensure the convergence of updates for  $\mathbf{A}$ ,  $\mathbf{S}$ ,  $\mathbf{G}$ , and  $\alpha^{(v)}$ , respectively. For steps 5 and 8, which update  $\mathbf{C}$  and  $\mathbf{F}$ , the objective values of Eqs. (9) and (14) are guaranteed to decrease monotonically, as demonstrated in [31]. Furthermore, the convergence behavior of Algorithm 1 will be empirically validated in the experimental section.

## IV. EXPERIMENTS

In this section, comprehensive experiments are conducted on multiple real-world datasets to evaluate the effectiveness

TABLE I  
A DETAILED DESCRIPTION OF DATASETS

Datasets	Abbr.	# Instances	# Classes	# Views	# Features
Yale	Yale	165	15	3	4096/3304/6750
MSRA	MSRA	210	7	6	1302/48/512/100/256/210
Politics-ie	Poli	348	7	9	14377/1047/1051/348/ 348/348/348/348/348
Mfeat	Mfeat	2000	10	6	216/76/64/6/240/47
Scene-15	Scene	4485	15	3	20/59/40
USPS-MNIST	USPS	9298	10	2	256/32
Alois	Alois	11025	100	4	77/13/64/64
SensIT Vehicle	Sens	15000	3	2	50/50

of SHINE-FS in comparison with state-of-the-art approaches.

### A. Experimental Schemes

1) *Datasets*: In this study, we use eight real-world multi-view datasets in our experiments, which are described as follows:

**Yale** [33]: 165 grayscale facial images from 15 individuals, with 11 images per subject captured under various lighting conditions and facial expressions.

**MSRA** [34]: 210 object images divided into seven categories, each represented by six distinct views: CENT, CMT, GIST, HOG, LBP, and SIFT.

**Politics-ie** [35]: 348 Twitter users divided into seven groups based on political affiliation, with each user characterized by nine views covering both network and textual information.

**Mfeat** [32]: 2,000 handwritten digit samples, each described by six feature sets extracted using different descriptors.

**Scene-15** [36]: 4,485 scene images across 15 indoor and outdoor categories, with each image represented by three views: GIST, PHOG, and LBP.

**USPS-MNIST** [37]: 9,298 handwritten digit samples belonging to 10 classes, with each sample obtained from two sources: USPS and MNIST.

**Alois** [38]: 11,025 images of 100 unique objects, photographed under diverse illumination and viewing angles, with each image represented by four feature sets: RGB, HSV, color similarity, and Haralick descriptors.

**SensIT Vehicle** [39]: 30,000 instances collected via a wireless sensor network, with each instance capturing signals from acoustic and seismic sensors for vehicle classification. Following [40], a subset of 15,000 samples is constructed by randomly selecting 5,000 samples from each class.

A detailed description of these multi-view datasets is provided in Table I.

2) *Evaluation Metrics*: Following [22], [41], we adopt two commonly used metrics in MUFS to assess the quality of the selected features, namely Clustering Accuracy (ACC) and Normalized Mutual Information (NMI). Their definitions are as follows:

$$\text{ACC} = \frac{1}{n} \sum_{i=1}^n \delta(y_i, \text{map}(y'_i)), \quad (17)$$

where  $n$  is the total number of samples, and  $y_i$  and  $y'_i$  correspond to the ground-truth and predicted cluster labels of the  $i$ -th sample, respectively. The function  $\text{map}(\cdot)$  matches

each cluster to its optimal true label using the Kuhn–Munkres algorithm [42], and  $\delta(x, y)$  is an indicator function that equals 1 when  $x = y$  and 0 otherwise.

$$\text{NMI} = \frac{\text{MI}(\mathcal{C}, \mathcal{C}')}{\max(H(\mathcal{C}), H(\mathcal{C}'))}, \quad (18)$$

where  $\mathcal{C}$  and  $\mathcal{C}'$  denote the sets of ground-truth and predicted clusters, respectively.  $\text{MI}(\mathcal{C}, \mathcal{C}')$  is the mutual information between  $\mathcal{C}$  and  $\mathcal{C}'$ , and  $H(\mathcal{C})$  and  $H(\mathcal{C}')$  are their corresponding entropies. For both metrics above, higher values indicate better performance.

3) *Compared Methods*: To evaluate the effectiveness of SHINE-FS, we compare it with several state-of-the-art (SOTA) methods, including three single-view methods (LPScore, HSL, and GAWFS) and nine multi-view methods (GCDUFS, UKMFS, CFSMO, EMUFS, CDMvFS, JMVFG, SDFS, CCSFS, and TLR-MUFS). Below is a brief introduction to these methods.

- **Allfea** utilizes all features from different views for comparison.
- **LPScore** [9] performs feature selection using the Laplacian score, which assesses the importance of each feature based on its ability to preserve local data structure.
- **HSL** [10] integrates feature selection with the learning of both first-order and higher-order similarities within a unified framework.
- **GAWFS** [11] simultaneously performs clustering and feature selection by combining non-negative matrix factorization with adaptive graph learning.
- **GCDUFS** [13] identifies important features by leveraging both consensus information across views and inter-view diversity.
- **UKMFS** [14] proposes a unified feature selection framework that incorporates robust self-representation and binary hashing.
- **CFSMO** [18] utilizes multi-order neighborhood information from similarity graphs, together with consistency across views, to enhance feature selection.
- **EMUFS** [22] jointly learns cluster assignments and local data structures by integrating membership matrices and bipartite graphs to identify discriminative features.
- **CDMvFS** [23] combines similarity graph learning with spectral clustering to select informative features.
- **JMVFG** [12] integrates orthogonal decomposition and consensus graph learning to preserve local data structure for feature selection.
- **SDFS** [20] simultaneously learns view-specific and consensus similarity graphs using an automatic weighting strategy to preserve data locality and guide feature selection.
- **CCSFS** [24] unifies subspace learning, consensus cluster discovery, and feature selection within a single framework.
- **TLR-MUFS** [21] identifies important features by capturing intra-view structures from view-specific graphs and enforcing consistency across views via tensor low-rank regularization.

TABLE II

MEANS AND STANDARD DEVIATIONS (%) OF ACC FOR VARIOUS METHODS ACROSS DIFFERENT DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Datasets Methods \	Yale	MSRA	Poli	Mfeat	Scene	USPS	Aloi	Sens
SHINE-FS	<b>57.01 ± 4.69</b>	<b>92.70 ± 1.26</b>	<b>58.02 ± 4.38</b>	<b>93.11 ± 1.71</b>	<b>41.96 ± 0.75</b>	<b>75.32 ± 0.51</b>	<b>59.45 ± 1.84</b>	<b>62.04 ± 0.00</b>
AllFea	49.41 ± 4.34•	51.19 ± 3.40•	40.80 ± 1.84•	50.31 ± 1.21•	29.45 ± 0.77•	35.50 ± 0.34•	16.76 ± 0.20•	60.45 ± 0.01•
LPScore	43.70 ± 3.47•	52.43 ± 0.46•	40.57 ± 1.59•	46.93 ± 0.16•	28.68 ± 0.22•	35.46 ± 0.24•	16.71 ± 0.23•	59.85 ± 0.01•
HSL	48.42 ± 2.83•	51.14 ± 3.27•	40.37 ± 1.34•	49.04 ± 0.82•	28.84 ± 1.03•	35.48 ± 0.22•	16.72 ± 0.20•	60.00 ± 0.00•
GAWFS	50.36 ± 3.77•	62.11 ± 2.95•	41.62 ± 2.56•	41.61 ± 0.95•	25.88 ± 0.81•	24.52 ± 0.03•	16.55 ± 0.74•	44.69 ± 0.00•
GCDUFS	56.30 ± 4.50◦	79.13 ± 3.54•	42.33 ± 3.39•	59.17 ± 2.42•	36.37 ± 0.39•	32.15 ± 0.78•	45.12 ± 1.37•	49.04 ± 0.42•
UKMFS	54.51 ± 3.68•	87.48 ± 1.19•	50.76 ± 6.29•	66.89 ± 5.66•	31.79 ± 0.57•	60.75 ± 0.24•	52.96 ± 1.42•	58.49 ± 0.00•
CFSMO	45.78 ± 3.02•	51.81 ± 2.09•	40.89 ± 1.94•	67.22 ± 0.38•	28.18 ± 0.35•	35.56 ± 0.38•	46.00 ± 1.71•	59.88 ± 0.00•
EUMFS	49.33 ± 4.57•	58.06 ± 1.22•	42.03 ± 2.26•	69.41 ± 0.94•	32.44 ± 0.15•	54.38 ± 0.49•	51.66 ± 1.29•	52.20 ± 0.23•
CDMvFS	47.86 ± 3.17•	53.14 ± 1.11•	40.78 ± 1.21•	70.27 ± 2.32•	37.80 ± 0.42•	35.63 ± 0.03•	34.03 ± 0.77•	59.81 ± 0.00•
JMVFG	55.74 ± 4.83◦	78.92 ± 1.43◦	40.02 ± 3.61•	72.78 ± 0.70•	35.93 ± 0.26•	58.71 ± 0.14•	46.89 ± 1.25•	47.67 ± 0.02•
SDFS	56.97 ± 4.29◦	31.92 ± 1.61•	41.89 ± 7.00•	74.21 ± 4.00•	36.60 ± 0.92•	59.96 ± 0.13•	48.76 ± 1.50•	60.00 ± 0.00•
CCSFS	47.88 ± 3.48•	62.94 ± 5.35•	40.80 ± 1.85•	46.20 ± 1.14•	27.66 ± 0.49•	36.24 ± 0.10•	53.19 ± 1.86•	60.00 ± 0.00•
TLR-MUFS	47.92 ± 4.49•	51.22 ± 3.52•	40.95 ± 1.21•	39.13 ± 0.24•	34.95 ± 0.48•	35.59 ± 0.04•	16.68 ± 0.28•	43.55 ± 0.00•

TABLE III

MEANS AND STANDARD DEVIATIONS (%) OF NMI FOR VARIOUS METHODS ACROSS DIFFERENT DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Datasets Methods \	Yale	MSRA	Poli	Mfeat	Scene	USPS	Aloi	Sens
SHINE-FS	<b>62.92 ± 3.02</b>	<b>86.48 ± 1.73</b>	<b>54.71 ± 3.12</b>	<b>86.62 ± 0.97</b>	<b>40.49 ± 0.44</b>	<b>60.23 ± 0.33</b>	<b>77.45 ± 0.73</b>	<b>18.59 ± 0.00</b>
AllFea	55.48 ± 2.64•	41.16 ± 0.86◦	7.57 ± 2.18•	56.80 ± 0.87•	28.72 ± 0.30•	25.96 ± 0.18•	37.33 ± 0.12•	16.74 ± 0.01•
LPScore	50.70 ± 2.21•	41.38 ± 0.40•	8.12 ± 2.77•	53.72 ± 0.11•	25.54 ± 0.16•	25.97 ± 0.14•	37.31 ± 0.12•	15.97 ± 0.01•
HSL	53.82 ± 1.89•	40.99 ± 1.22•	6.69 ± 1.99•	55.90 ± 0.64•	27.03 ± 0.24•	25.97 ± 0.14•	37.29 ± 0.14•	16.18 ± 0.00•
GAWFS	55.75 ± 2.72•	51.00 ± 2.27•	12.69 ± 3.14•	29.53 ± 0.61•	25.50 ± 0.41•	14.30 ± 0.10•	32.27 ± 1.27•	7.11 ± 0.00•
GCDUFS	61.03 ± 3.66•	70.02 ± 2.06•	31.24 ± 3.68•	63.17 ± 1.35•	37.95 ± 0.40•	23.57 ± 0.63•	66.31 ± 0.75•	9.38 ± 0.41•
UKMFS	60.64 ± 3.48•	79.14 ± 1.72•	48.15 ± 5.08•	71.08 ± 3.41•	34.59 ± 0.36•	51.24 ± 0.17•	73.67 ± 0.62•	14.21 ± 0.00•
CFSMO	52.00 ± 2.18•	41.12 ± 0.47•	7.93 ± 2.82•	62.39 ± 0.36•	26.05 ± 0.18•	25.92 ± 0.21•	69.78 ± 0.71•	16.04 ± 0.00•
EUMFS	54.67 ± 3.61•	47.96 ± 1.17•	11.21 ± 3.45•	66.51 ± 0.74•	30.22 ± 0.11•	51.76 ± 0.14•	72.02 ± 0.53•	9.93 ± 0.09•
CDMvFS	53.23 ± 2.56•	43.42 ± 1.21•	6.89 ± 2.26•	66.79 ± 1.61•	36.16 ± 0.51•	26.48 ± 0.02•	56.58 ± 0.37•	15.94 ± 0.00•
JMVFG	61.08 ± 4.37•	69.74 ± 1.71•	8.94 ± 3.00•	63.95 ± 0.65•	37.65 ± 0.25•	52.92 ± 0.08•	66.43 ± 0.55•	8.82 ± 0.02•
SDFS	60.41 ± 4.27•	17.01 ± 2.06•	15.01 ± 4.58•	72.17 ± 1.71•	34.46 ± 0.59•	51.60 ± 0.08•	70.46 ± 0.72•	16.17 ± 0.00•
CCSFS	53.53 ± 2.55•	55.90 ± 2.49•	7.93 ± 2.72•	41.70 ± 0.42•	28.08 ± 0.41•	26.51 ± 0.02•	74.44 ± 0.69•	16.17 ± 0.00•
TLR-MUFS	53.37 ± 3.28•	45.21 ± 2.12•	7.40 ± 2.05•	46.68 ± 0.10•	38.17 ± 0.32•	25.89 ± 0.01•	37.29 ± 0.14•	5.14 ± 0.00•

To ensure a fair comparison, we use grid search to tune the parameters of all methods and report their best performance. In our method, the parameters  $\gamma$ ,  $\beta$ , and  $\eta$  are tuned using the same strategy within the range  $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$ . Since selecting the optimal number of features for a given dataset remains challenging [1], [43], we vary the percentage of selected features from 10% to 50% in increments of 10% for all datasets. Then, k-means clustering [44] is performed 30 times on the selected features, and the average results are reported. All experiments were carried out using MATLAB R2022b on a desktop with an Intel Core i9-10900 CPU (2.80 GHz) and 64 GB of RAM.

### B. Performance Comparison

In this section, we assess the clustering performance of SHINE-FS in comparison with other methods using ACC and NMI metrics. Tables II and III show the ACC and NMI results on eight datasets with a feature selection ratio of 20%. To further demonstrate the superiority of SHINE-FS, we use the Wilcoxon rank-sum test [45] to determine whether its performance is significantly better than that of the

competing approaches. In Tables II and III, values marked with • indicate that SHINE-FS performs significantly better than the compared methods at the 0.05 significance level, while values marked with ◦ indicate that SHINE-FS does not show a significant improvement over the compared methods.

As shown in Tables II and III, the proposed method consistently outperforms the compared approaches. As to MSRA, Mfeat, and USPS datasets, SHINE-FS achieves average improvements of more than 34% in ACC and 31% in NMI. For Poli and Aroi datasets, SHINE-FS achieves average improvements exceeding 20% in ACC and 31% in NMI. For Yale, Scene, and Sens datasets, SHINE-FS also outperforms the competing methods, with average increases of over 7% in both ACC and NMI.

Furthermore, since it is difficult to determine the optimal number of selected features, we compare the performance of various methods under different feature selection ratios in terms of ACC and NMI. The ACC and NMI results are illustrated in Figs. 2 and 3, respectively. As observed in these figures, the proposed method achieves superior performance over the competing methods in most cases. The superior

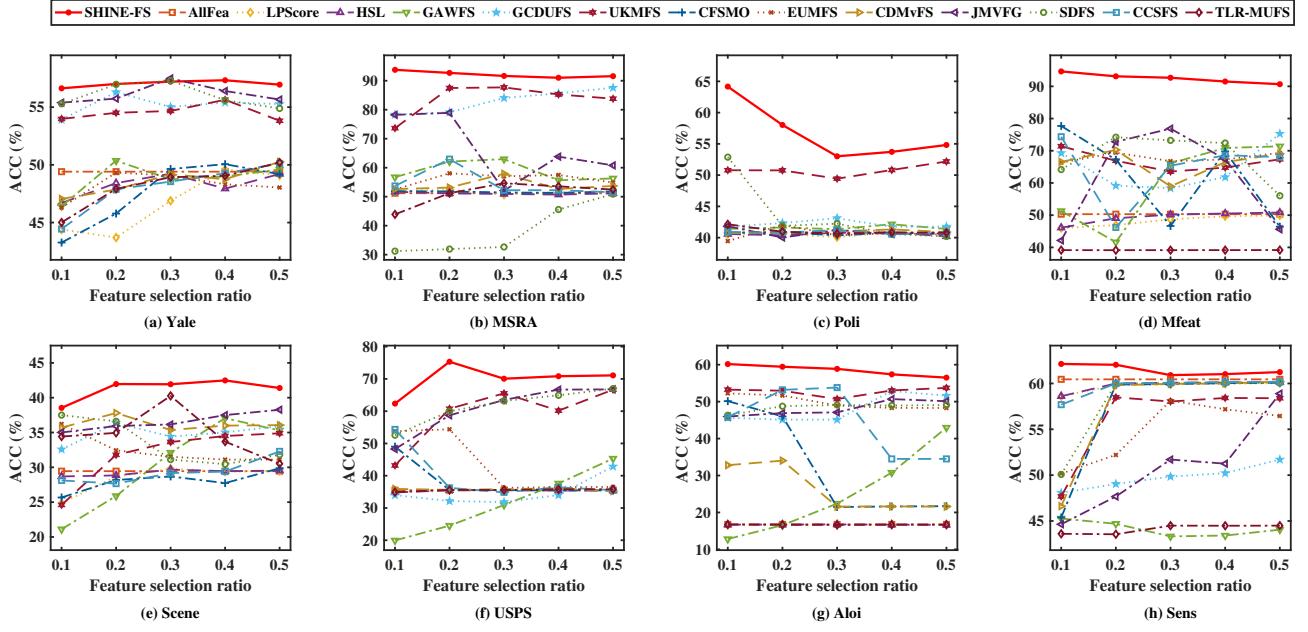


Fig. 2. ACC of different methods on eight datasets with different feature selection ratios.

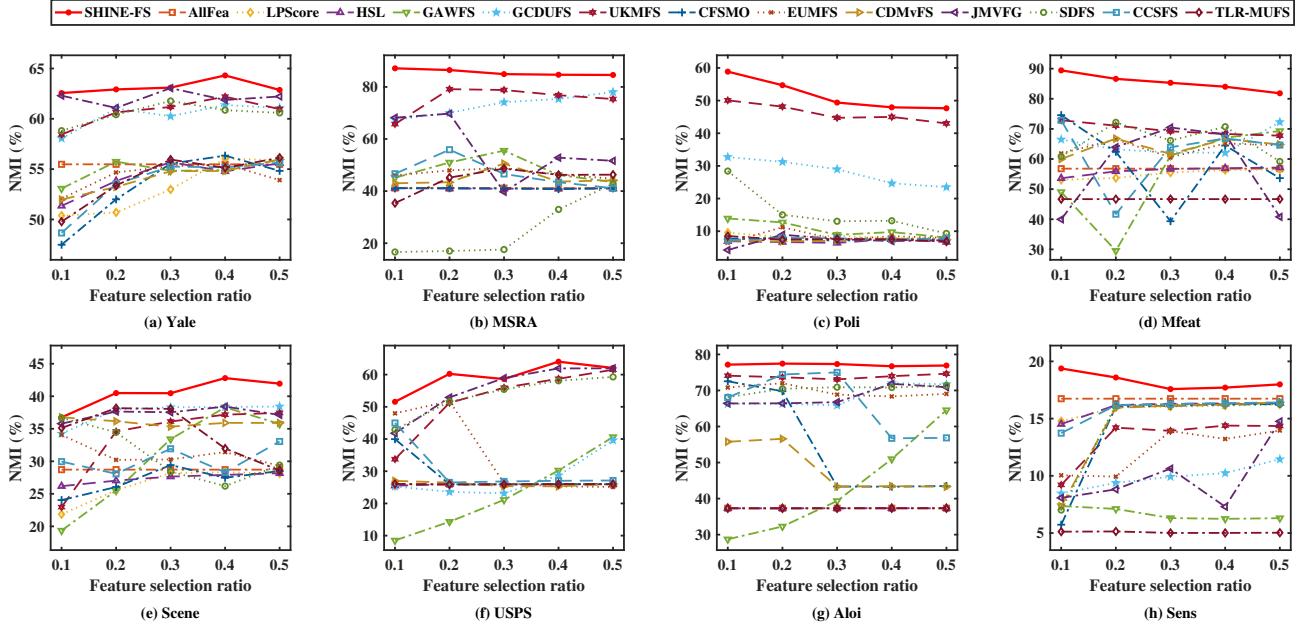


Fig. 3. NMI of different methods on eight datasets with different feature selection ratios.

performance of SHINE-FS can be attributed to its unified framework, which seamlessly combines multi-view feature selection, consensus anchor graph learning, and hybrid-order similarity learning. This integration enables the model to capture both local and global structures, providing a more effective characterization of the intrinsic data structure, thus better guiding feature selection.

### C. Visualization Analysis

To intuitively demonstrate the effectiveness of the learned hybrid-order similarity graph in capturing the intrinsic data

structure, we visualize the similarity graphs generated by different graph-based methods on MSRA dataset, as shown in Fig. 4. It can be observed that the hybrid-order similarity graph learned by SHINE-FS (Fig. 4 (e)) exhibits a clearer structure that is highly consistent with the ground-truth graph (Fig. 4 (f)). In contrast, the graphs generated by other methods (Figs. 4 (a)-(d)) show ambiguous structures that poorly represent the sample relationships. These results demonstrate the superiority of SHINE-FS in uncovering the intrinsic data structure.

Furthermore, to visually evaluate the quality of the selected

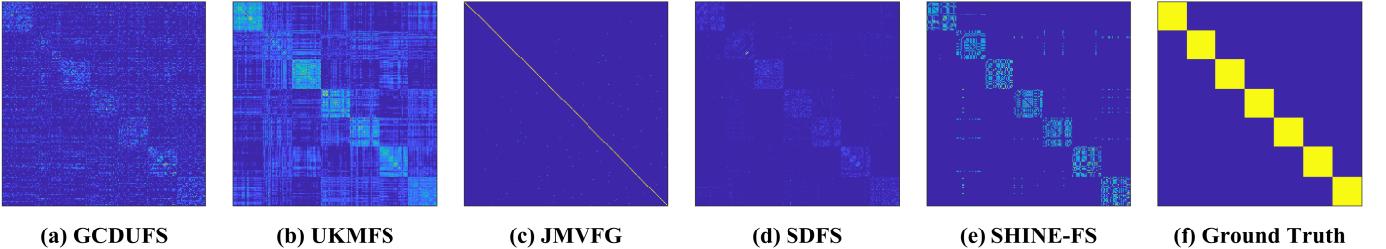


Fig. 4. Visualizations of the similarity graphs learned by different graph-based methods on MSRA dataset.

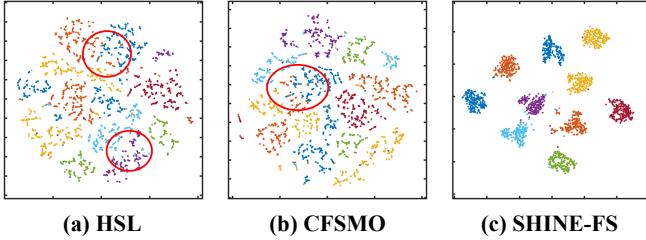


Fig. 5. The t-SNE visualizations on Mfeat dataset.

features, we use t-SNE [46] to embed the samples represented by these features into a two-dimensional space. Fig. 5 illustrates the t-SNE visualization results on Mfeat dataset with a 20% feature selection ratio. Figs. 5 (a) and (b) show the results of HSL and CFSMO, respectively, both of which rely on predefined high-order similarities, while Fig. 5 (c) presents the visualization result of SHINE-FS. The results indicate that HSL and CFSMO show considerable overlap among clusters, while SHINE-FS produces greater inter-cluster separation and clearer boundaries. This demonstrates that SHINE-FS can effectively select a compact subset of discriminative features by adaptively learning a hybrid-order similarity graph.

#### D. Parameter Sensitivity and Convergence Behavior

The objective function in Eq. (3) includes five parameters:  $\gamma$ ,  $\beta$ ,  $\eta$ ,  $\lambda_S$ , and  $\lambda_G$ . Among these,  $\lambda_S$  and  $\lambda_G$  are automatically determined when solving for  $\mathbf{S}$  in Eq. (11) and  $\mathbf{G}$  in Eq. (13), respectively. To investigate the sensitivity of SHINE-FS to the parameters  $\gamma$ ,  $\beta$ , and  $\eta$ , Figs. 6 and 7 illustrate how its ACC and NMI performance changes as these parameters and the feature selection ratio (FR) are varied on the MSRA dataset. As illustrated in these figures, SHINE-FS remains relatively stable for  $\gamma$  and  $\beta$ , while slight fluctuations are observed as  $\eta$  varies. In practice, the parameter  $\eta$  should be tuned using a grid search to achieve optimal performance.

Furthermore, to verify the convergence of Algorithm 1, we present its convergence behavior curves on MSRA, Mfeat, and Scene datasets. As shown in Fig. 8, the objective value of SHINE-FS drops sharply during the initial iterations and stabilizes rapidly, typically within the first 10 iterations. This demonstrates that our method converges quickly.

#### E. Ablation Study

To demonstrate the effectiveness of the proposed module in SHINE-FS, we conduct ablation experiments by comparing it

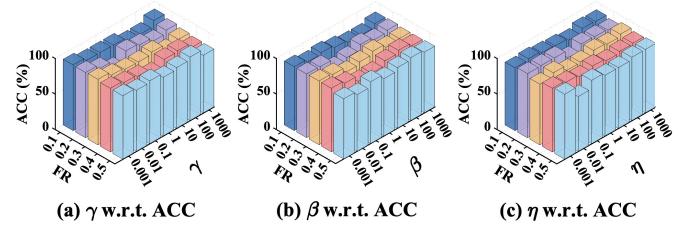
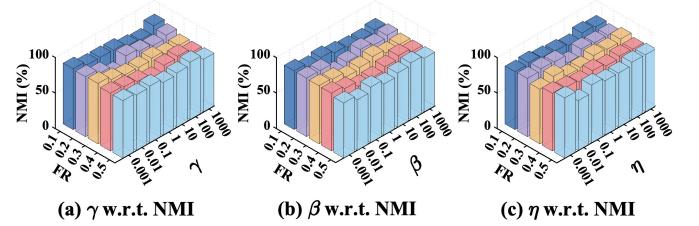
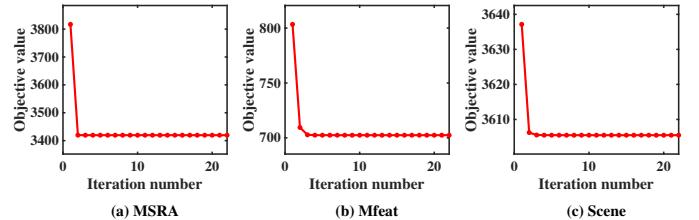
Fig. 6. ACC of SHINE-FS with varying parameters  $\gamma$ ,  $\beta$ ,  $\eta$  and feature selection ratios on MSRA dataset.Fig. 7. NMI of SHINE-FS with varying parameters  $\gamma$ ,  $\beta$ ,  $\eta$  and feature selection ratios on MSRA dataset.

Fig. 8. Convergence curves of SHINE-FS on MSRA, Mfeat and Scene datasets.

with three of its variants. Specifically, SHINE-FS-I excludes only the anchor-induced second-order similarity graph learning module. SHINE-FS-II removes the hybrid-order similarity graph learning module. SHINE-FS-III replaces the consensus anchor graph learning module with a view-specific anchor graph learning module. Fig. 9 presents the ablation experiment results in terms of ACC and NMI on eight datasets. It can be observed that: (i) SHINE-FS consistently outperforms SHINE-FS-I, validating that the anchor-induced second-order similarity graph can effectively capture the global structure of the data. (ii) The superior performance of SHINE-FS over SHINE-FS-II demonstrates the effectiveness of adaptive hybrid-order similarity graph learning for jointly capturing local and global structures, thus enabling a more comprehensive characteriza-

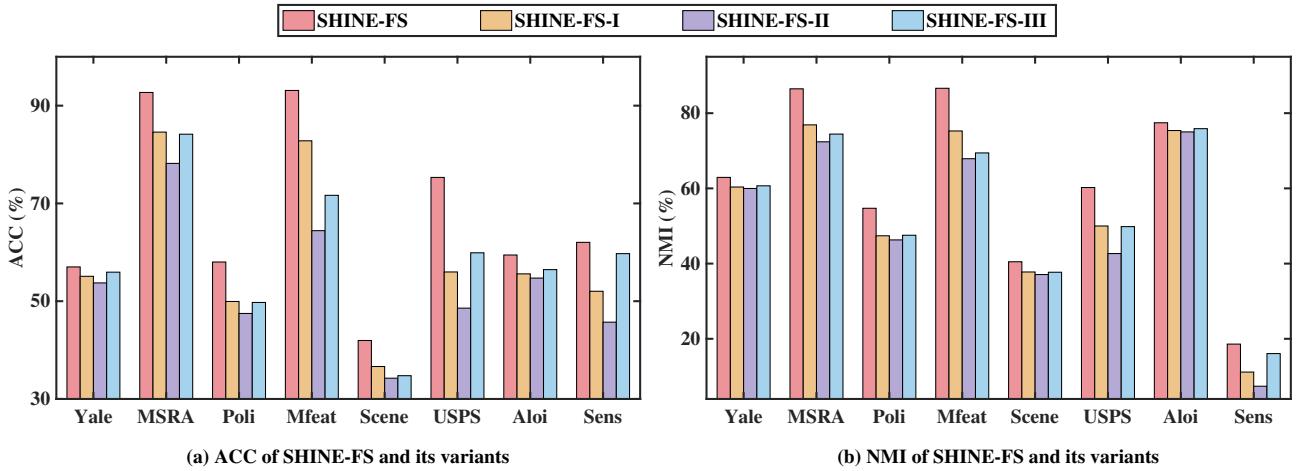


Fig. 9. Performance comparison of SHINE-FS and its three variants in terms of ACC and NMI.

tion of the intrinsic data structure. (iii) SHINE-FS outperforms SHINE-FS-III, indicating that consensus anchor graph learning exploits shared structural information across views to facilitate feature selection.

## V. CONCLUSION

In this paper, we propose a novel MUFS method called SHINE-FS to address the challenge of capturing the intrinsic data structure in MUFS. SHINE-FS first learns consensus anchors and the corresponding anchor graph to capture anchor-sample relationships across views, which facilitates the learning of a second-order similarity graph. Furthermore, SHINE-FS jointly learns the first-order and second-order similarity graphs to construct a hybrid similarity graph that effectively integrates local and global structural information, thus better characterizing the intrinsic data structure. Extensive experiments on real multi-view datasets demonstrate the superiority of SHINE-FS over SOTA methods. In many practical applications, labeled samples are often scarce; however, they provide essential discriminative information that can effectively guide feature selection. Therefore, our future work will aim to extend the proposed unsupervised framework to a semi-supervised scenario, enabling the identification of informative features from multi-view data with limited labels available.

## REFERENCES

- [1] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, “Feature selection: A data perspective,” *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, 2017.
- [2] B. Pan, C. Li, H. Che, M.-F. Leung, and K. Yu, “Low-rank tensor regularized graph fuzzy learning for multi-view data processing,” *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 2925–2938, 2023.
- [3] R. Zhang, F. Nie, X. Li, and X. Wei, “Feature selection with multi-view data: A survey,” *Information Fusion*, vol. 50, pp. 158–167, 2019.
- [4] Q. Y. Qiang, B. Zhang, C. J. Zhang, and F. P. Nie, “Adaptive bipartite-based multi-view unsupervised dimensionality reduction,” *Neural Networks*, vol. 188, p. 107424, 2025.
- [5] C. Hou, F. Nie, H. Tao, and D. Yi, “Multi-view unsupervised feature selection with adaptive similarity and view weight,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1998–2011, 2017.
- [6] B. Jiang, X. Wu, X. Zhou, Y. Liu, A. G. Cohn, W. Sheng, and H. Chen, “Semi-supervised multiview feature selection with adaptive graph learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 3, pp. 3615–3629, 2022.
- [7] X. Yang, H. Che, M.-F. Leung, and S. Wen, “Self-paced regularized adaptive multi-view unsupervised feature selection,” *Neural Networks*, vol. 175, p. 106295, 2024.
- [8] Z. W. Cao and X. J. Xie, “Structure learning with consensus label information for multi-view unsupervised feature selection,” *Expert Systems with Applications*, vol. 238, p. 121893, 2024.
- [9] X. He, D. Cai, and P. Niyogi, “Laplacian score for feature selection,” in *Proceedings of the 19th Advances in Neural Information Processing Systems*, 2005, pp. 507–514.
- [10] Y. Mi, H. Chen, C. Luo, S.-J. Horng, and T. Li, “Unsupervised feature selection with high-order similarity learning,” *Knowledge-Based Systems*, vol. 285, p. 1111317, 2024.
- [11] H. Liao, H. Chen, T. Yin, Z. Yuan, S.-J. Horng, and T. Li, “A general adaptive unsupervised feature selection with auto-weighting,” *Neural Networks*, vol. 181, p. 106840, 2025.
- [12] S. Fang, D. Huang, C. Wang, and Y. Tang, “Joint multi-view unsupervised feature selection and graph learning,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 1, pp. 16–31, 2024.
- [13] S. Xu, X. Xie, and Z. Cao, “Graph-regularized consensus learning and diversity representation for unsupervised multi-view feature selection,” *Knowledge-Based Systems*, vol. 311, p. 113043, 2025.
- [14] R. Hu, J. Gan, M. Zhan, L. Li, and M. Wei, “Unsupervised kernel-based multi-view feature selection with robust self-representation and binary hashing,” in *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, 2025, pp. 17287–17294.
- [15] C. Zhang, X. Liang, P. Zhou, Z. Ling, Y. Zhang, X. Wu, W. Sheng, and B. Jiang, “Scalable multi-view unsupervised feature selection with structure learning and fusion,” in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 5479–5488.
- [16] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “Line: Large-scale information network embedding,” in *Proceedings of the 24th International Conference on World Wide Web*, 2015, p. 1067–1077.
- [17] B. Thijs, E. Schiebel, and W. Glänzel, “Do second-order similarities provide added-value in a hybrid approach?” *Scientometrics*, vol. 96, no. 3, pp. 667–677, 2013.
- [18] Z. Cao and X. Xie, “Multi-view unsupervised complementary feature selection with multi-order similarity learning,” *Knowledge-Based Systems*, vol. 283, p. 111172, 2024.
- [19] K. Jiang, Z. Yang, and Q. Sun, “Block-diagonal graph embedding for unsupervised feature selection,” *Applied Intelligence*, vol. 55, no. 7, p. 674, 2025.
- [20] S. Zhou, P. Song, Y. Yu, and W. Zheng, “Structural regularization based discriminative multi-view unsupervised feature selection,” *Knowledge-Based Systems*, vol. 272, p. 110601, 2023.
- [21] H. Yuan, J. Li, Y. Liang, and Y. Y. Tang, “Multi-view unsupervised feature selection with tensor low-rank minimization,” *Neurocomputing*, vol. 487, pp. 75–85, 2022.

- [22] C. Zhang, Y. Fang, X. Liang, X. Wu, and B. Jiang, "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.
- [23] Z. Cao, X. Xie, and Y. Li, "Multi-view unsupervised feature selection with consensus partition and diverse graph," *Information Sciences*, vol. 661, p. 120178, 2024.
- [24] Z. Cao, X. Xie, F. Sun, and J. Qian, "Consensus cluster structure guided multi-view unsupervised feature selection," *Knowledge-Based Systems*, vol. 271, p. 110578, 2023.
- [25] S. X. Zhou and P. Song, "Consistency-exclusivity guided unsupervised multi-view feature selection," *Neurocomputing*, vol. 569, p. 127119, 2024.
- [26] C. Liang, L. Z. Wang, L. Liu, H. X. Zhang, and F. Guo, "Multi-view unsupervised feature selection with tensor robust principal component analysis and consensus graph learning," *Pattern Recognition*, vol. 141, p. 109632, 2023.
- [27] R. Wang, P. Wang, D. Wu, Z. Sun, F. Nie, and X. Li, "Multi-view and multi-order structured graph learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 10, pp. 14437–14448, 2023.
- [28] F. Nie, R. Zhang, and X. Li, "A generalized power iteration method for solving quadratic problems on the stiefel manifold," *Science China Information Sciences*, vol. 60, pp. 1–10, 2017.
- [29] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering," in *Proceedings of the 24th International Conference on Artificial Intelligence*, 2015, pp. 3569–3575.
- [30] S. Wang, X. Liu, E. Zhu, C. Tang, J. Liu, J. Hu, J. Xia, and J. Yin, "Multi-view clustering via late fusion alignment maximization," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019, pp. 3778–3784.
- [31] F. Nie, W. Zhu, and X. Li, "Unsupervised feature selection with structured graph optimization," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 2016, pp. 1302–1308.
- [32] H. Wang, Y. Yang, and B. Liu, "Gmc: Graph-based multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1116–1129, 2019.
- [33] Y. Sun, X. Li, Q. Sun, M.-L. Zhang, and Z. Ren, "Improved weighted tensor schatten  $p$ -norm for fast multi-view graph clustering," in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 1427–1436.
- [34] X. Yu, H. Liu, Y. Zhang, Y. Lin, and C. Zhang, "Hubness-enabled clustering and recovery for large-scale incomplete multi-view data," *ACM Transactions on Knowledge Discovery from Data*, vol. 19, no. 1, pp. 1–23, 2024.
- [35] Q. Zhou, W. Zhu, H. Chen, and B. Peng, "Community detection in multiplex networks by deep structure-preserving non-negative matrix factorization," *Applied Intelligence*, vol. 55, no. 1, p. 26, 2025.
- [36] D. Li, H. Wang, and S. Ying, "Multiview representation learning with one-to-many dynamic relationships," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 7, pp. 13051–13065, 2024.
- [37] J. J. Hull, "A database for handwritten text recognition research," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 550–554, 2002.
- [38] A. Rocha and S. K. Goldenstein, "Multiclass from binary: Expanding one-versus-all, one-versus-one and ecoc-based approaches," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 289–302, 2013.
- [39] X. Cai, F. Nie, and H. Huang, "Multi-view k-means clustering on big data," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2013, pp. 2598–2604.
- [40] L. Zhao, T. Yang, J. Zhang, Z. Chen, Y. Yang, and Z. J. Wang, "Co-learning non-negative correlated and uncorrelated features for multi-view data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1486–1496, 2020.
- [41] Y. Y. Huang, Z. X. Shen, T. R. Li, and F. M. Lv, "UNIFIER: Unified view imputation and feature selection learning for incomplete multi-view data," in *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 2024, pp. 4192–4200.
- [42] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *Journal of Machine Learning Research*, vol. 3, no. 12, pp. 583–617, 2002.
- [43] Z. W. Cao and X. J. Xie, "Partition-level tensor learning-based multiview unsupervised feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 7, pp. 12799–12811, 2024.
- [44] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1965, pp. 281–298.
- [45] R. F. Woolson, "Wilcoxon signed-rank test," *Wiley Encyclopedia of Clinical Trials*, vol. 8, pp. 1–3, 2007.
- [46] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2579–2605, 2008.