Semi-Supervised Multi-View Clustering based on NMF with Fusion Regularization

GUOSHENG CUI, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China and Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen, China
RUXIN WANG, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China and Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen, China
DAN WU, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China and Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen, China
YE LI, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China and Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen, China

Multi-view clustering has attracted significant attention and application. Nonnegative matrix factorization is one popular feature of learning technology in pattern recognition. In recent years, many semi-supervised nonnegative matrix factorization algorithms were proposed by considering label information, which has achieved outstanding performance for multi-view clustering. However, most of these existing methods have either failed to consider discriminative information effectively or included too much hyper-parameters. Addressing these issues, a semi-supervised multi-view nonnegative matrix factorization with a novel fusion regularization (FRSMNMF) is developed in this article. In this work, we uniformly constrain alignment of multiple views and discriminative information among clusters with designed fusion regularization. Meanwhile, to align the multiple views effectively, two kinds of compensating matrices are used to normalize the feature scales of different views. Additionally, we preserve the geometry structure information of labeled and unlabeled samples by introducing the graph regularization simultaneously. Due to the proposed methods, two effective optimization strategies based on multiplicative update rules are designed. Experiments implemented on six real-world datasets have demonstrated the effectiveness of our FRSMNMF comparing with several state-of-the-art unsupervised and semi-supervised approaches.

CCS Concepts: • **Theory of computation → Unsupervised learning and clustering;**

This work was supported in part by National Key Research and Development Program of China under Grant 2022YFA1008300, in part by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant XDB38040200, in part by the National Natural Science Foundation of China under Grants 62206269, 62073310 and 62102410, in part by the Joint Funds of the National Natural Science Foundation of China under Grants U2241210 and U1913210, in part by Guangdong Provincial Science and Technology Plan under Grant 2022A1515011217 and 2022A1515011557, in part by Shenzhen Outstanding Youth Project under Grant RCYX20200714114641211.
Authors' address: G. Cui, R. Wang, D. Wu, and Y. Li, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China, 518055 and Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen, Guangdong, China, 518055; e-mails: gs.cui@siat.ac.cn, rx.wang@siat.ac.cn, dan.wu@siat.ac.cn, ye.li@siat.ac.cn.

ACM 1556-4681/2024/04-ART157
https://doi.org/10.1145/3653022

Additional Key Words and Phrases: Multi-view, semi-supervised clustering, fusion regularization, scale normalization

ACM Reference Format:
Guosheng Cui, Ruxin Wang, Dan Wu, and Ye Li. 2024. Semi-Supervised Multi-View Clustering based on NMF with Fusion Regularization. *ACM Trans. Knowl. Discov. Data.* 18, 6, Article 157 (April 2024), 26 pages. https://doi.org/10.1145/3653022

1 INTRODUCTION

Clustering is a very important multi-variant analysis technology in pattern recognition and machine learning. In past decades, many clustering methods [1, 21, 50, 63] have been proposed such as *k*-means [15], Gaussian mixture model, spectral clustering, hierarchical clustering algorithm, and so on. While most of them are designed for single view data. However, in many cases, the data in the real world are collected from multiple sources [18, 34, 55], which characterizes the objects from various perspectives and is usually termed as multiple views or modalities in these literatures [9, 47, 51, 61]. The different views generally contain interaction and complementary information which can be integrated for enhancing the recognition performance for different tasks. Therefore, for multiview clustering tasks, how to effectively fuse the information from the multiple views to enhance the clustering performance is the most important issue in the learning process [2, 59].

Among various outstanding multiview clustering methods [7, 8, 45, 48, 56, 57], **nonnegative matrix factorization (NMF)** [24, 25] based methods are widely studied owing to the excellent feature learning ability of NMF. Under the framework of **multi-view nonnegative matrix factorization (MVNMF)** [35, 38, 64], each point can be represented with an efficient low dimensional feature vector. From the perspective of whether to utilize label information, MVNMF methods [32] can be divided into unsupervised methods [46] and semi-supervised methods [28]. Many unsupervised MVNMFs focus on exploring the fusion mechanism of multiple views in hidden space. For example, in [46], a novel pair-wise co-regularization was proposed to align the multiple views, which can consider the similarity of the inter-view. In [14], a nonredundancy regularization was developed to discover the distinct contributions of the different views. In [42] and [27], the inter-view diversity terms are developed to learn more comprehensive information of data. Additionally, unsupervised MVNMFs usually take some normalizing operations to improve the performance of the model. For instance, in [41], to ensure the numerical stablility of NMF, in each iteration the column summation of the basis matrix is normalized to be 1, i.e., $\sum_i \mathbf{W}^v_{i\cdot} = 1$ (v denotes the vth view). Furthermore, to make the representations of the different views comparable at the same scale, in [52], [39], and [32], the same normalization operation in [41] is adopted and the norms of the basis matrix are compensated into the coefficient matrix in the co-regularization term. To guarantee the solution uniqueness of NMF, in [17], the row-summation of the coefficient matrix is normalized to be 1, i.e., $\sum_j \mathbf{H}^v_{\cdot j} = 1$, where each column denotes a data point. These methods have failed to consider the label information that the data may provide. In fact, effectively utilizing the label information provided by a small proportion of the data, the clustering performance can be further enhanced. Under this consideration, in recent years, some semi-supervised MVNMFs are developed trying to employ the label information.

Most recently proposed semi-supervised MVNMF methods are based on **constrained NMF (CNMF)** [30] which is a semi-supervised NMF method for single view data. The shortage of CNMF-based semi-supervised MVNMFs [5, 6, 43, 54] is that they all have failed to discover the discriminative information among the different classes. In these methods, the data points of the same classes

are presented by the same feature vectors. By this means, the distances among the data points residing in the same classes are zeroed. However, the distances among the data points residing in the different classes are not maximized. Another problem of these CNMF-based methods is that they all have failed to make use of the local geometrical information of the data. In [22], Jiang et al. have tried to discover the discriminative information of data by regressing the partial coefficient matrix to the label matrix. The work in [31] has also considered the discriminative information of data. Unfortunately, both of these methods have ignored the geometrical information of data. The defect of these methods has been overcome in the work of [28] which has extended the work of [31] by including a graph regularization. Liang et al. [29] have tried to make use of the geometrical information of data by involving an adaptive local structure learning module. In [58], hypergraph regularization is used to encode the structure of data.

Although, the methods in [28], [29], and [58] have considered both discriminative information and geometrical information of the data. The price is that too many hyper-parameters make these methods hard to get satisfying performance for different datasets. Another problem is that, normalizing strategy is widely disccussed in unsupervised MVNMF methods; however, it is rarely explored under the framework of semi-supervised MVNMF. Addressing these issues, in this article, a **semi-supervised multi-view nonnegative matrix factorization with fusion regularization (FRSMNMF)** is proposed. In this method, the discriminative term and the feature alignment term are fused as one regularization prior termed as fusion regularization. The meaning of this regularization comes from two aspects. On the one hand, the features of different views are fused by this regularization. On the other hand, the feature fusion constraint and discriminative constraint are integrated as one constraint with this regularization, which effectively reduces the number of hyper-parameters. Additionally, to further make use of geometrical information of the data, graph regularization is also constructed for each view. To align the feature scale of different views, two different feature normalizing strategies are adopted, which corresponds to two variants denoted as FRSMNMF_N1 and FRSMNMF_N2. And the influences of these two normalizing strategies are compared in the experiments. For the two variants, corresponding specific iterative optimizing schemes are also designed. The contributions of this article are summarized as follows:

— A novel fusion regularization based semi-supervised MVNMF is presented. In this work, the exploration of discriminative information and feature alignment are achieved in one term which reduces the number of hyper-parameters and enlarges the inter-class distinction;
— Two feature normalizing strategies are adopted to align the feature scales of different views, which produces two variants of the proposed framework;
— For the two variants of the proposed framework, two specific iterative optimizing schemes are designed to solve the corresponding minimization problems effectively;
— The effectiveness of the proposed methods is evaluated by comparing with some recently proposed representative unsupervised and semi-supervised MVNMFs on six datasets.

The rest of this article is organized as follows: Section 2 introduces some related works on multi-view clustering. In Section 3, the details of the proposed semi-supervised multi-view nonnegative matrix factorization with fusion regularization (FRSMNMF) and its two specific variants are introduced. Section 4 gives the corresponding optimizing strategy, convergence proof and computational complexity analysis. In Section 5, the experimental results are demonstrated. Finally, the conclusion of this article is made in Section 6.

2 RELATED WORK

In this section, we briefly review the related unsupervised and semi-supervised multi-view clustering methods.

2.1 Unsupervised Multi-view Clustering

Most of unsupervised MVNMF methods are based on the idea of aligning the multiple views in a shared common subspace to fuse the information. In [32], Liu et al. proposed a pioneer MVNMF method which tries to learn a consensus representation by minimizing a centroid co-regularization. With this regularization, the feature matrices of the different views are pushed toward a consensus feature matrix, and then multiview alignment is achieved. Furthermore, a feature scale normalization matrix is introduced to constrain the feature scales of different views to be similar hoping to facilitate the feature alignment. In [39], Rai et al. have adopted the same strategy to make the scales of different views comparable. In [52], Yang et al. have explored another way to reduce the distribution divergences of the feature matrix. They presented a MVNMF based on nonnegative matrix tri-factorization [53]. In order to restrict the feature scale of different views to be similar, the column-summation of the product matrix of basis matrix and shared embedding matrix are constrained to be 1. Essentially, Liu et al. [32], Rai et al. [39], and Yang et al. [52] have adopted the same feature normalizing strategy, in which the column-summation of the basis matrix is constrained to be 1, i.e., $\sum_i \mathbf{W}_{i\cdot}^v = 1$, and the norms are compensated to the coefficient matrix in the co-regularization term. Actually, to improve the performance of unsupervised MVNMFs, various normalizing operations are explored. For example, to guarantee the numerical stability of NMF, Shao et al. [41] have constrained the column-summation of the basis matrix to be 1 and do not compensate the norms to the coefficient matrix in the co-regularization term. To ensure the solution uniqueness of NMF, Hu et al. [17] have constrained the row-summation of the coefficient matrix to be 1, i.e., $\sum_j \mathbf{H}_{\cdot j}^v = 1$. Most of existing methods are based on the centroid co-regularization. While, in [46], a pair-wise co-regularization was proposed to align the multiple views pair-wisely. By minimizing this co-regularization, the alignment is acquired through pushing the representations of different views together. Until now, we can find that, most of the MVNMF methods were developed from aligning the multiple views, and ignored the diversity of the different views. Addressing this issue, in [42], an MVNMF called *diverse nonnegative matrix factorization* has been proposed. In this method, a diverse term is designed to encourage the heterogeneity of different views and try to learn some more comprehensive information. Recently, with the flourish of deep learning in computer vision and pattern recognition, some "deep" models are proposed for multi-view clustering task [13]. For example, a deep multi-view semi-nonnegative matrix factorization is proposed in [60]. In this method, an adaptive weighting strategy is adopted to balance the effect of different views. Similarly, a deep MVNMF is proposed in [26]. The graph regularization is applied in all layers not only in the final layer like [60] and all factor matrices are restricted to be nonnegative. All of the previously reviewed methods focus on dealing with "well-established" data. It means that all views have the corrected correspondence on sample-level. In [20], Huang et al. have developed a partially view-aligned clustering method which can tackle the multi-view data not fully view-aligned. "Partially view-aligned" means that only a part of data have the corrected correspondence on sample-level.

2.2 Semi-Supervised Multi-View Clustering

In recent years, several semi-supervised MVNMFs are proposed and promising results are obtained with label information. Wang et al. [43] have proposed a semi-supervised MVNMF based on CNMF for the first time. In this work, an adaptive weighting strategy is applied to balance the effect of different views. To learn a consensus representation, a similar centroid co-regularization like [32] is utilized in their method. In [5] and [6], Cai et al. have developed two semi-supervised MVNMF methods; these two methods are also based on CNMF. Both methods have adopted pair-wise co-regularization with Euclidean distance to align multiple views like [46]. The difference between [5]

and [6] is the second regularizing term. In [5], $\ell_{2,1}$-norm regularization of the auxiliary matrix is used, and an orthonormality constraint on the auxiliary matrix is used in [6]. The former one tries to get sparse representations for each view, and the latter one tries to constrain the feature scale of different views to be similar. Yang et al. [54] have improved the work in [6] by relaxing the label constraint matrix. The common drawback of above methods is that they all have ignored the discriminative information of data. To make use of this information, Jiang et al. [22] have proposed a unified latent factor learning model trying to reconstruct the label matrix with the product of partial shared coefficient matrix and an auxiliary matrix. Further, Liu et al. [31] have improved the above work by splitting the coefficient matrix into the private part and the common part. Based on [31], Liang et al. [28] have tried to improve the performance of the method by imposing a graph regularization on the coefficient matrix. The problem with [28] is that it uses too many hyper-parameters, which makes this method very complex and hard to fine-tune. Wang et al. [44] have presented a semi-supervised multi-view clustering model with weighted anchor graph embedding, in which the anchors are constructed based on label information. Nie et al. [36] have developed a semi-supervised multi-view learning method which can model the structure of data adaptively. Based on this work, Liang et al. [29] have presented a semi-supervised MVNMF method based on label propagation. Zhang et al. [58] have proposed a semi-supervised MVNMF method, called *dual hypergraph regularized partially shared NMF*, in which hypergraph regularization was imposed on both the coefficient matrix and the basis matrix. Similar to [28], these two methods also suffer from involving too many hyper-parameters. Recently, some deep multi-view clustering methods were also developed. Zhao et al. [62] have proposed a deep semi-supervised MVNMF method which encodes the discriminative information of data by constructing an affinity graph for intra-class compactness and a penalty graph for inter-class distinctness. Chen et al. [11] have presented an autoencoder-based semi-supervised multi-view clustering method which encodes the discriminative information of data by introducing a pairwise constraint.

From above review, we can find that, although various matrix normalizing strategies are widely explored in many unsupervised MVNMF methods, these stategies are rarely discussed under the semi-supervised MVNMF framework. In these previously introduced normalizing strategies, the column-summation of the basis matrix is usually normalized to be 1, and sometimes the norms are compensated to the coefficient matrix in the co-regularization term. However, it is more reasonable to constrain the columns of the basis to be unit vectors, i.e., $||\mathbf{W}^v_{:j}||_2 = 1$. Additionally, recently proposed semi-supervised MVNMF methods have involved too many hyper-parameters, which has made these methods very complex and hard to fine-tune. Addressing these issues, in the following sections, we will introduce a novel semi-supervised MVNMF framework with fusion regularization. This framework can be implemented with two different feature normalizing strategies. And the fusion regularization can make use of the discriminative information of data and reduce the number of hyper-parameters.

3 SEMI-SUPERVISED MULTI-VIEW CLUSTERING WITH FUSION REGULARIZATION

For a traditional MVNMF framework, the objective function can be written as follows:

$$\sum_v ||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \lambda\mathcal{R}(\mathbf{H}^v), \tag{1}$$

where $\mathbf{X}^v \in \mathbb{R}_+^{m^v \times n}$, $\mathbf{W}^v \in \mathbb{R}_+^{m^v \times d^v}$ and $\mathbf{H}^v \in \mathbb{R}_+^{d^v \times n}$ are the data matrix, the basis matrix and the coefficient matrix of the vth view, respectively. m^v and d^v are the dimensions of original space and latent space for the vth view, and n is the size of the dataset. $\mathcal{R}(\mathbf{H}^v)$ denotes the regularization term which is employed to introduce extra prior or constrains for enhancing the clustering performance, such as manifold regularization, co-regularization. λ is the hyper-parameter which

regulates the participation of these constraints in the optimization process. Both the selection of hyper-parameter λ and regularization prior $\mathcal{R}(\mathbf{H}^v)$ play the important role in the design of the model. In order to achieve satisfactory clustering result, enlarge the inter-class distinction, shrink the distance of intra-class, and achieve multi-view feature alignment effectively are crucial. However, most of the existing multi-view NMF methods have failed to consider discriminative information and feature alignment simultaneously. Although some algorithms like AMVNMF, MVCNMF, and MVOCNMF have adopted label information of the data, they have not further considered the distinguishability among classes. Therefore, performance of clustering is limited. In addition, the model should avoid introducing too many hyper-parameters while integrating more regular priors $\mathcal{R}(\mathbf{H}^v)$ for feature learning.

For enforcing the inter-class distinctness with multi-view features, in this article, we first construct a novel fusion regularization which considers the distinct information among clusters and alignment of different views together. It can be formulated as follows:

$$||[\mathbf{Y}, \mathbf{H}_c] - [\mathbf{H}_l^v, \mathbf{H}_{ul}^v]||_F^2 \tag{2}$$

where $\mathbf{Y} \in \{0,1\}^{d^v \times n_l}$ is the label matrix and $\mathbf{H}_l^v \in \mathbb{R}_+^{d^v \times n_l}$ is the feature matrix of labeled samples for the vth view. n_l is the number of labeled data points. $\mathbf{H}_c \in \mathbb{R}_+^{d^v \times n_{ul}}$ denotes the common feature matrix and $\mathbf{H}_{ul}^v \in \mathbb{R}_+^{d^v \times n_{ul}}$ is the feature matrix of unlabeled data points for the vth view. n_{ul} is the number of unlabeled data points. Obviously, in this work, $d^1 = d^2 = \cdots = d^V = C$. C is the number of classes. A general way to simultaneously consider discriminative information and feature alignment is to define them separately as follows:

$$||\mathbf{Y} - \mathbf{H}_l^v||_F^2 + ||\mathbf{H}_c^* - \mathbf{H}^v||_F^2, \tag{3}$$

where $\mathbf{H}^v = [\mathbf{H}_l^v, \mathbf{H}_{ul}^v]$ and $\mathbf{H}_c^* = [\mathbf{H}_{cl}^*, \mathbf{H}_c]$. Then, Equation (3) can be further written as follows:

$$||\mathbf{Y} - \mathbf{H}_l^v||_F^2 + ||[\mathbf{H}_{cl}^*, \mathbf{H}_c] - [\mathbf{H}_l^v, \mathbf{H}_{ul}^v]||_F^2. \tag{4}$$

In Equation (4), the first term is used to align the coefficient matrix of the labeled samples (i.e., $\mathbf{H}_l^v$) to the label matrix (i.e., $\mathbf{Y}$), and the second term is used to align the whole coefficient matrix of all samples (i.e., $\mathbf{H}^v = [\mathbf{H}_l^v, \mathbf{H}_{ul}^v]$) to the common consensus matrix $\mathbf{H}_c^* = [\mathbf{H}_{cl}^*, \mathbf{H}_c]$. We can see that the representation of the labeled samples ($\mathbf{H}_l^v$) is aligned by both $\mathbf{Y}$ and $\mathbf{H}_{cl}^*$. The goal of aligning $\mathbf{H}_l^v$ to $\mathbf{Y}$ is to discover the discriminative information of data. While, the goal of aligning $\mathbf{H}_l^v$ to $\mathbf{H}_{cl}^*$ is to align the multiple views to push the representations of different views ($\mathbf{H}_l^v$, $v = 1, 2, \ldots, V$) close to each other. Actually, by aligning $\mathbf{H}_l^v$ to $\mathbf{Y}$ can also achieve this goal, because the label matrix $\mathbf{Y}$ is shared by all views. With this intuition, we substitute $\mathbf{H}_{cl}^*$ with $\mathbf{Y}$ in Equation (4) and remove the first term, then, we can get the proposed Equation (2). Until now, we can see that by aligning $\mathbf{H}_l^v$ with $\mathbf{Y}$, the discriminative information can be discovered and $\mathbf{H}_l^v$ ($v = 1, 2, \ldots, V$) of different views can be pushed close to each other. And $\mathbf{H}_{ul}^v$ ($v = 1, 2, \ldots, V$) of different views can be pushed together by aligning them to $\mathbf{H}_c$. The meaning of fusion regularization has two folds: (1) the features of different views are fused by this regularization; (2) the feature fusion constraint and discriminative constraint are integrated as one constraint with this regularization, which reduces the introduction of hyper-parameters.

In Equations (2), (3), and (4), the label matrix $\mathbf{Y}$ is approximated by $\mathbf{H}_l^v$. We can denote the data matrix $\mathbf{X}^v$ as $\mathbf{X}^v = [\mathbf{X}_l^v, \mathbf{X}_{ul}^v]$ corresponding to the low dimensional representation $\mathbf{H}^v = [\mathbf{H}_l^v, \mathbf{H}_{ul}^v]$. Then $\mathbf{H}_l^v$ is the low dimensional representation of $\mathbf{X}_l^v$. Actually, the label matrix $\mathbf{Y}$ can also be seen as a feature representation of $\mathbf{X}_l^v$, which contains the discriminative information of data. So, by subtracting $\mathbf{Y}$ with $\mathbf{H}_l^v$, i.e., minimizing $||\mathbf{Y} - \mathbf{H}_l^v||_F^2$, $\mathbf{Y}$ is reconstructed by $\mathbf{H}_l^v$, then the

discriminative information in $\mathbf{Y}$ can be learned by $\mathbf{H}_l^v$. Note that, we do not constrain the column summation of $\mathbf{H}_l^v$ to be 1, i.e., $\sum_i \mathrm{H}_{l(i)}^v \neq 1$, so $\mathbf{H}_l^v$ is a relaxed representation of $\mathbf{Y}$.

Additionally, to make use of the geometry information of the multi-view data in each view for decreasing the intra-class diversities, a graph regularizer is constructed based on both labeled samples and unlabeled samples, which is defined as follows:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|h_i^v - h_j^v\right\|_2^2 \mathbf{S}_{ij}^v = \mathrm{tr}(\mathbf{H}^{vT}\mathbf{L}^v\mathbf{H}^v), \tag{5}$$

where $\mathbf{L}^v = \mathbf{D}^v - \mathbf{S}^v$, $\mathbf{D}_{ii}^v = \sum_j \mathbf{S}_{ij}^v$ (or $\mathbf{D}_{ii}^v = \sum_j \mathbf{S}_{ji}^v$). $\mathbf{S}^v$ is defined as follows:

$$\mathbf{S}_{ij}^v = \begin{cases} 1 & \text{if } h_i^v \in N_k(h_j^v) \text{ or } h_j^v \in N_k(h_i^v) \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

where $N_k(h_i^v)$ consists of k nearest neighbors of h_i^v.

Considering the above properties of inter- and intra-class, the overall objective function of our proposed **semi-supervised multi-view nonnegative matrix factorization with fusion regularization (FRSMNMF)** can be described as follows:

$$\begin{aligned} O_{\mathrm{FRSMNMF}} = \sum_{v=1}^{V} \{&||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \alpha\mathrm{tr}(\mathbf{H}^{vT}\mathbf{L}^v\mathbf{H}^v) \\ &+ \beta||[\mathbf{Y}, \mathbf{H}_c] - [\mathbf{H}_l^v, \mathbf{H}_{ul}^v]||_F^2\} \\ s.t.\ \mathbf{W}^v, \mathbf{H}^v, \mathbf{H}_c \geq 0.& \end{aligned} \tag{7}$$

To tackle multi-view tasks, another important issue is to align multiple features effectively for aggregating the information of different views. Instead of optimizing Equation (7) directly, we first restrict the scales of multiple features to be comparable-level. For this target, the column vectors of $\mathbf{W}^v$ are constrained be $||\mathbf{W}_{\cdot j}^v||_2 = 1$ or $||\mathbf{W}_{\cdot j}^v||_1 = 1$. One possible way to account for the above constraints is to add an extra regularization term to Equation (7), but this would make the optimization problem very complicated. Instead of taking the direct strategy, an alternative scheme is to compensate the norms of the basis matrix into the coefficient matrix. Then, the objective function of FRSMNMF in Equation (7) can be rewritten as:

$$\begin{aligned} O_{\mathrm{FRSMNMF}} = \sum_{v=1}^{V} \{&||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \alpha\mathrm{tr}(\mathbf{Q}^v\mathbf{H}^v\mathbf{L}^v\mathbf{H}^{vT}\mathbf{Q}^{vT}) \\ &+ \beta||[\mathbf{Y}, \mathbf{H}_c] - \mathbf{Q}^v\mathbf{H}^v||_F^2\} \\ s.t.\ \mathbf{W}^v, \mathbf{H}^v, \mathbf{H}_c \geq 0,& \end{aligned} \tag{8}$$

where $\mathbf{Q}^v$ is defined as

$$\mathbf{Q}^v = Diag\left(\sqrt{\sum_{i=1}^{m^v} {\mathbf{W}_{i,1}^v}^2}, \sqrt{\sum_{i=1}^{m^v} {\mathbf{W}_{i,2}^v}^2}, \ldots, \sqrt{\sum_{i=1}^{m^v} {\mathbf{W}_{i,d^v}^v}^2}\right), \tag{9}$$

or

$$\mathbf{Q}^v = Diag\left(\sum_{i=1}^{m^v} \mathbf{W}_{i,1}^v, \sum_{i=1}^{m^v} \mathbf{W}_{i,2}^v, \ldots, \sum_{i=1}^{m^v} \mathbf{W}_{i,d^v}^v\right). \tag{10}$$

For simplicity, the proposed methods with $||\mathbf{W}_{\cdot j}^v||_2 = 1$ and $||\mathbf{W}_{\cdot j}^v||_1 = 1$ are denoted as FRSMNMF_N1 and FRSMNMF_N2, respectively, corresponding to two different normalization manners. In following sections, the optimization algorithms of FRSMNMF_N1 and FRSMNMF_N2 will be introduced in detail.

4 OPTIMIZATION PROCEDURE

4.1 Optimization Algorithm of FRSMNMF_N1

The optimization problem of minimizing Equation (8) is not direct. In this section, an alternating optimization strategy is developed, which breaks the original problem into several subproblems such that each subproblem is tractable.

For a specific view v, its optimization with $\mathbf{W}^v$ and $\mathbf{H}^v$ does not depend on other views. The problem of minimizing Equation (8) can be written as follows:

$$\begin{aligned}\min_{\mathbf{W}^v,\mathbf{H}^v,\mathbf{H}_c\geq 0} &||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \alpha \mathrm{tr}(\mathbf{Q}^v\mathbf{H}^v\mathbf{L}^v\mathbf{H}^{vT}\mathbf{Q}^{vT}) \\ &+ \beta||[\mathbf{Y},\mathbf{H}_c] - \mathbf{Q}^v\mathbf{H}^v||_F^2\end{aligned} \tag{11}$$

4.1.1 Fixing $\boldsymbol{H}_c$, updating $\boldsymbol{W}^v$ and $\boldsymbol{H}^v$. (1) **Fixing $\mathbf{H}_c$ and $\mathbf{H}^v$, updating $\mathbf{W}^v$**

When $\mathbf{H}_c$ and $\mathbf{H}^v$ are fixed, the above equation is equivalent to the following problem:

$$\begin{aligned}\min_{\mathbf{W}^v\geq 0} &||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \alpha \mathrm{tr}(\mathbf{Q}^v\mathbf{H}^v\mathbf{L}^v\mathbf{H}^{vT}\mathbf{Q}^{vT}) \\ &+ \beta \mathrm{tr}(\mathbf{Q}^v\mathbf{H}^v\mathbf{H}^{vT}\mathbf{Q}^{vT} - 2\mathbf{H}^v\mathbf{H}_{yc}^T\mathbf{Q}^{vT}),\end{aligned} \tag{12}$$

where $\mathbf{H}_{yc} = [\mathbf{Y},\mathbf{H}_c]$.

Let

$$\begin{aligned}\mathbf{Y}_1^v &= [\mathbf{H}^v\mathbf{L}^v\mathbf{H}^{vT}]\odot\mathbf{I} \\ &= \mathbf{Y}_1^{v+} - \mathbf{Y}_1^{v-} \\ \mathbf{Y}_2^v &= [\mathbf{H}^v\mathbf{H}^{vT}]\odot\mathbf{I},\end{aligned} \tag{13}$$

where $\odot$ is the element-wise product operator, $\mathbf{I}$ is the identity matrix and the matrices $\mathbf{Y}_1^{v+}$ and $\mathbf{Y}_1^{v-}$ are defined as

$$\begin{aligned}\mathbf{Y}_1^{v+} &= [\mathbf{H}^v\mathbf{D}^v\mathbf{H}^{vT}]\odot\mathbf{I} \\ \mathbf{Y}_1^{v-} &= [\mathbf{H}^v\mathbf{S}^v\mathbf{H}^{vT}]\odot\mathbf{I}.\end{aligned} \tag{14}$$

Then, Equation (12) with $\mathbf{Q}^v$ defined as in Equation (9) can be further rewritten as

$$\begin{aligned}\min_{\mathbf{W}^v\geq 0} &||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \alpha \mathrm{tr}(\mathbf{W}^v\mathbf{Y}_1^v\mathbf{W}^{vT}) \\ &+ \beta \mathrm{tr}(\mathbf{W}^v\mathbf{Y}_2^v\mathbf{W}^{vT} - 2\mathbf{H}^v\mathbf{H}_{yc}^T\mathbf{Q}^{vT}).\end{aligned} \tag{15}$$

As $\mathbf{W}^v \geq 0$, introducing the Lagrangian multipliers ψ_{ih} for the constraints $\mathbf{W}_{ih}^v \geq 0$. Let $\Psi = [\psi_{ih}]$. Then, the Lagrangian function $\mathcal{L}$ is as follows:

$$\begin{aligned}\mathcal{L} = &||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \alpha \mathrm{tr}(\mathbf{W}^v\mathbf{Y}_1^v\mathbf{W}^{vT}) \\ &+ \beta \mathrm{tr}(\mathbf{W}^v\mathbf{Y}_2^v\mathbf{W}^{vT} - 2\mathbf{H}^v\mathbf{H}_{yc}^T\mathbf{Q}^{vT}) + \mathrm{tr}(\Psi\mathbf{W}^{vT})\end{aligned} \tag{16}$$

The partial derivative of $\mathrm{tr}(\mathbf{H}^v\mathbf{H}_{yc}^T\mathbf{Q}^{vT})$ with respect to $\mathbf{W}_{ih}^v$ is

$$\begin{aligned}\frac{\partial \mathrm{tr}(\mathbf{H}^v\mathbf{H}_{yc}^T\mathbf{Q}^{vT})}{\partial \mathbf{W}_{ih}^v} &= \frac{\partial (\mathbf{H}^v\mathbf{H}_{yc}^T)_{hh}\sqrt{\sum_{k=1}^{m^v}\mathbf{W}_{kh}^{v\ 2}}}{\partial \mathbf{W}_{ih}^v} \\ &= (\mathbf{H}^v\mathbf{H}_{yc}^T)_{hh}\frac{\mathbf{W}_{ih}^v}{\sqrt{\sum_{k=1}^{m^v}\mathbf{W}_{kh}^{v\ 2}}} \\ &= (\mathbf{W}^v(\mathbf{Q}^v)^{-1}(\mathbf{H}^v\mathbf{H}_{yc}^T\odot\mathbf{I}))_{ih} \\ &= (\mathbf{W}^v(\mathbf{Q}^v)^{-1}\mathbf{Y}_3^v)_{ih},\end{aligned} \tag{17}$$

where $\mathbf{Y}_3^v = [\mathbf{H}^v\mathbf{H}_{yc}^T]\odot\mathbf{I}$.

Until now, the partial derivative of (16) with respect to $\mathbf{W}^v$ can be written as follows:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{W}^v} &= 2\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT} - 2\mathbf{X}^v\mathbf{H}^{vT} + 2\alpha\mathbf{W}^v\mathbf{Y}_1^v \\ &\quad + 2\beta\mathbf{W}^v\mathbf{Y}_2^v - 2\beta\mathbf{W}^v(\mathbf{Q}^v)^{-1}\mathbf{Y}_3^v + \Psi.\end{aligned} \tag{18}$$

Setting above equation to zero and using the **Karush–Kuhn–Tucker (KKT) condition** [3] of $\psi_{ih}\mathbf{W}_{ih}^v = 0$, then have

$$\begin{aligned}&2(\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT})_{ih}\mathbf{W}_{ih}^v - 2(\mathbf{X}^v\mathbf{H}^{vT})_{ih}\mathbf{W}_{ih}^v \\ &+ 2\alpha(\mathbf{W}^v\mathbf{Y}_1^v)_{ih}\mathbf{W}_{ih}^v + 2\beta(\mathbf{W}^v\mathbf{Y}_2^v)_{ih}\mathbf{W}_{ih}^v \\ &- 2\beta(\mathbf{W}^v(\mathbf{Q}^v)^{-1}\mathbf{Y}_3^v)_{ih}\mathbf{W}_{ih}^v \\ &= 0,\end{aligned} \tag{19}$$

which leads to the following update rule:

$$\mathbf{W}_{ih}^v = \mathbf{W}_{ih}^v \frac{(\mathbf{X}^v\mathbf{H}^{vT} + \alpha\mathbf{W}^v\mathbf{Y}_1^{v-} + \beta\mathbf{W}^v(\mathbf{Q}^v)^{-1}\mathbf{Y}_3^v)_{ih}}{(\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT} + \alpha\mathbf{W}^v\mathbf{Y}_1^{v+} + \beta\mathbf{W}^v\mathbf{Y}_2^v)_{ih}} \tag{20}$$

(2) **Fixing $\mathbf{H}_c$ and $\mathbf{W}^v$, updating $\mathbf{H}^v$**

After the updating of $\mathbf{W}^v$, the column vectors of $\mathbf{W}^v$ are normalized with $\mathbf{Q}^v$ in Equation (9) and then the norm is conveyed to the coefficient matrix $\mathbf{H}^v$, that is:

$$\mathbf{W}^v \Leftarrow \mathbf{W}^v(\mathbf{Q}^v)^{-1}, \mathbf{H}^v \Leftarrow \mathbf{Q}^v\mathbf{H}^v. \tag{21}$$

When $\mathbf{H}_c$ and $\mathbf{W}^v$ are fixed, Equation (11) is equivalent to the following problem:

$$\begin{aligned}\min_{\mathbf{H}^v \geq 0} &||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \alpha\mathrm{tr}(\mathbf{H}^v\mathbf{L}^v\mathbf{H}^{vT}) \\ &+ \beta\mathrm{tr}(\mathbf{H}^v\mathbf{H}^{vT} - 2\mathbf{H}^v\mathbf{H}_{yc}^T).\end{aligned} \tag{22}$$

As $\mathbf{H}^v \geq 0$, introducing the Lagrangian multipliers ϕ_{jh} for the constraints $\mathbf{H}_{jh}^v \geq 0$. Let $\Phi = [\phi_{jh}]$. Then, the Lagrangian function $\mathcal{L}$ is as follows:

$$\begin{aligned}\mathcal{L} = &||\mathbf{X}^v - \mathbf{W}^v\mathbf{H}^v||_F^2 + \alpha\mathrm{tr}(\mathbf{H}^v\mathbf{L}^v\mathbf{H}^{vT}) \\ &+ \beta\mathrm{tr}(\mathbf{H}^v\mathbf{H}^{vT} - 2\mathbf{H}^v\mathbf{H}_{yc}^T) + \mathrm{tr}(\Phi\mathbf{H}^{vT}).\end{aligned} \tag{23}$$

The partial derivative of (23) with respect to $\mathbf{H}^v$ is as follows:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{H}^v} &= 2\mathbf{W}^{vT}\mathbf{W}^v\mathbf{H}^v - 2\mathbf{W}^{vT}\mathbf{X}^v \\ &\quad + 2\alpha\mathbf{H}^v\mathbf{D}^v - 2\alpha\mathbf{H}^v\mathbf{S}^v + 2\beta\mathbf{H}^v - 2\beta\mathbf{H}_{yc} + \Phi.\end{aligned} \tag{24}$$

Similarly, setting $\frac{\partial \mathcal{L}}{\partial \mathbf{H}^v} = 0$ and using KKT condition of $\phi_{hj}\mathbf{H}_{hj}^v = 0$, then we have

$$\begin{aligned}&2(\mathbf{W}^{vT}\mathbf{W}^v\mathbf{H}^v)_{hj}\mathbf{H}_{hj}^v - 2(\mathbf{W}^{vT}\mathbf{X}^v)_{hj}\mathbf{H}_{hj}^v + 2\alpha(\mathbf{H}^v\mathbf{D}^v)_{hj}\mathbf{H}_{hj}^v \\ &- 2\alpha(\mathbf{H}^v\mathbf{S}^v)_{hj}\mathbf{H}_{hj}^v + 2\beta(\mathbf{H}^v)_{hj}\mathbf{H}_{hj}^v - 2\beta(\mathbf{H}_{yc})_{hj}\mathbf{H}_{hj}^v \\ &= 0,\end{aligned} \tag{25}$$

which leads to the following update rules:

$$\mathbf{H}_{hj}^v = \mathbf{H}_{hj}^v \frac{(\mathbf{W}^{vT}\mathbf{X}^v + \alpha\mathbf{H}^v\mathbf{S}^v + \beta\mathbf{H}_{yc})_{hj}}{(\mathbf{W}^{vT}\mathbf{W}^v\mathbf{H}^v + \alpha\mathbf{H}^v\mathbf{D}^v + \beta\mathbf{H}^v)_{hj}}. \tag{26}$$

ALGORITHM 1: Optimizing scheme for FRSMNMF_N1

Input: Multi-view data $\mathbf{D}_X = \{\mathbf{X}^1, \ldots, \mathbf{X}^V\}$
Parameters $\alpha > 0$, $\beta > 0$, and the number of k nearest neighbors: K
Output: $\mathbf{H}^v$ and $\mathbf{H}_c$
for each $v \in V$ **do**
Initialize $\mathbf{W}^v \in \mathbb{R}_+^{m^v \times d^v}$, $\mathbf{H}^v \in \mathbb{R}_+^{d^v \times n}$, $\mathbf{H}_c \in \mathbb{R}_+^{d^v \times n_{ul}}$ with the random values from (0, 1] uniform distribution;
Construct k-NN graph with 0-1 weight;
end
repeat
for each $v \in V$ **do**
S1: Fix $\mathbf{H}^v$ and $\mathbf{H}_c$, update $\mathbf{W}^v$ with Equation (20);
S2: Normalize $\mathbf{W}^v$ and $\mathbf{H}^v$ with Equation (21),
while $\mathbf{Q}^v$ is defined as Equation (9);
S3: Fix $\mathbf{W}^v$ and $\mathbf{H}_c$, update $\mathbf{H}^v$ with Equation (26);
end
Fix $\mathbf{W}^v$ and $\mathbf{H}^v$, update $\mathbf{H}_c$ with Equation (28);
until convergence

4.1.2 Fixing $\boldsymbol{W}^v$ and $\boldsymbol{H}^v$, updating $\boldsymbol{H}_c$. As $\mathbf{W}^v$ is normalized in each iteration, the partial derivative of (8) with respect to $\mathbf{H}_c$ is as follows:

$$\begin{aligned}\frac{\partial O_{\text{FRSMNMF}}}{\partial \mathbf{H}_c} &= \frac{\partial \sum_{v=1}^{V} \beta ||\mathbf{H}_l^v - \mathbf{H}_c||_F^2}{\partial \mathbf{H}_c} \\ &= \sum_{v=1}^{V} [-2\beta \mathbf{H}_l^v + 2\beta \mathbf{H}_c] = 0.\end{aligned} \tag{27}$$

Then, the exact solution for $\mathbf{H}_c$ is

$$\mathbf{H}_c = \frac{\sum_{v=1}^{V} \mathbf{H}_l^v}{V} \geq 0. \tag{28}$$

The optimizing scheme of FRSMNMF_N1 is summarized in Algorithm 1.

4.2 Optimization Algorithm of FRSMNMF_N2

Comparing with FRSMNMF_N1, the main difference of optimizing procedure for FRSMNMF_N2 lies in the updating rule of $\mathbf{W}^v$. So, in this section, only the updating rule of $\mathbf{W}^v$ is deduced. Fixing $\mathbf{H}_c$ and $\mathbf{H}^v$, the objective function of FRSMNMF_N2 with respect to $\mathbf{W}^v$ is written as follows:

$$\begin{aligned}&\min_{\mathbf{W}^v \geq 0} ||\mathbf{X}^v - \mathbf{W}^v \mathbf{H}^v||_F^2 + \alpha \text{tr}(\mathbf{Q}^v \mathbf{H}^v \mathbf{L}^v \mathbf{H}^{vT} \mathbf{Q}^{vT}) \\ &+ \beta ||[\mathbf{Y}, \mathbf{H}_c] - \mathbf{Q}^v \mathbf{H}^v||_F^2,\end{aligned} \tag{29}$$

where $\mathbf{Q}^v$ is defined as in Equation (10).

As $\mathbf{W}^v \geq 0$, introducing the Lagrangian multipliers ψ_{ih} for the constraints $\mathbf{W}_{ih}^v \geq 0$. Let $\Psi = [\psi_{ih}]$. Then, the Lagrangian function $\mathcal{L}$ is as follows:

$$\begin{aligned}\mathcal{L} = &||\mathbf{X}^v - \mathbf{W}^v \mathbf{H}^v||_F^2 + \alpha \text{tr}(\mathbf{Q}^v \mathbf{H}^v \mathbf{L}^v \mathbf{H}^{vT} \mathbf{Q}^{vT}) \\ &+ \beta \text{tr}(\mathbf{Q}^v \mathbf{H}^v \mathbf{H}^{vT} \mathbf{Q}^{vT} - 2\mathbf{H}^v \mathbf{H}_{yc}{}^T \mathbf{Q}^{vT}) + \text{tr}(\Psi \mathbf{W}^{vT}).\end{aligned} \tag{30}$$

ALGORITHM 2: Optimizing scheme for FRSMNMF_N2

Input: Multi-view data $\mathbf{D}_X = \{\mathbf{X}^1, \ldots, \mathbf{X}^V\}$
Parameters $\alpha > 0$, $\beta > 0$, and the number of k nearest neighbors: K
Output: $\mathbf{H}^v$ and $\mathbf{H}_c$
for each $v \in V$ **do**
Initialize $\mathbf{W}^v \in \mathbb{R}_+^{m^v \times d^v}$, $\mathbf{H}^v \in \mathbb{R}_+^{d^v \times n}$, $\mathbf{H}_c \in \mathbb{R}_+^{d^v \times n_{ul}}$ with the random values from (0, 1] uniform distribution;
Construct k-NN graph with 0-1 weight;
end
repeat
for each $v \in V$ **do**
S1: Fix $\mathbf{H}^v$ and $\mathbf{H}_c$, update $\mathbf{W}^v$ with Equation (33);
S2: Normalize $\mathbf{W}^v$ and $\mathbf{H}^v$ with Equation (21),
while $\mathbf{Q}^v$ is defined as Equation (10);
S3: Fix $\mathbf{W}^v$ and $\mathbf{H}_c$, update $\mathbf{H}^v$ with Equation (26);
end
Fix $\mathbf{W}^v$ and $\mathbf{H}^v$, update $\mathbf{H}_c$ with Equation (28);
until convergence

The partial derivative of (30) with respect to $\mathbf{W}^v$ can be written as follows:

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{W}^v_{ih}} &= 2(\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT})_{ih} - 2(\mathbf{X}^v\mathbf{H}^{vT})_{ih} \\ &+ 2\alpha(\mathbf{Q}^v\mathbf{Y}^v_1)_{hh} + 2\beta(\mathbf{Q}^v\mathbf{Y}^v_2)_{hh} - 2\beta(\mathbf{Y}^v_3)_{hh} + \psi_{ih},\end{aligned} \tag{31}$$

where $\mathbf{Y}^v_1$, $\mathbf{Y}^v_2$ and $\mathbf{Y}^v_3$ are defined in Equation (13) and Equation (17).

Setting the above equation to zero and using the KKT condition of $\psi_{ih}\mathbf{W}^v_{ih} = 0$, then we have

$$\begin{aligned}&2(\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT})_{ih}\mathbf{W}^v_{ih} - 2(\mathbf{X}^v\mathbf{H}^{vT})_{ih}\mathbf{W}^v_{ih} \\ &+ 2\alpha(\mathbf{Q}^v\mathbf{Y}^v_1)_{hh} + 2\beta(\mathbf{Q}^v\mathbf{Y}^v_2)_{hh}\mathbf{W}^v_{ih} \\ &- 2\beta(\mathbf{Y}^v_3)_{hh} = 0,\end{aligned} \tag{32}$$

which leads to the following update rule:

$$\mathbf{W}^v_{ih} = \mathbf{W}^v_{ih}\frac{(\mathbf{X}^v\mathbf{H}^{vT})_{ih} + (\alpha\mathbf{Q}^v\mathbf{Y}^{v-}_1 + \beta\mathbf{Y}^v_3)_{hh}}{(\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT})_{ih} + (\alpha\mathbf{Q}^v\mathbf{Y}^{v+}_1 + \beta\mathbf{Q}^v\mathbf{Y}^v_2)_{hh}}. \tag{33}$$

The updating rules of $\mathbf{H}^v$ and $\mathbf{H}_c$ in FRSMNMF_N2 are the same as those in Equation (26) and Equation (28). The optimizing scheme of FRSMNMF_N2 is summarized in Algorithm 2. The convergence proofs for FRSMNMF_N1 and FRSMNMF_N2 are presented in the Appendix.

4.3 Computational Complexity Analysis of FRSMNMF_N1 and FRSMNMF_N2

In this section, the computational complexity of FRSMNMF_N1 and FRSMNMF_N2 is analyzed. The computational complexity is expressed in a big O notation [12]. For a specific view v of FRSMNMF_N1 in one iteration, updating $\mathbf{W}^v$ needs to calculate $\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT}$, $\mathbf{X}^v\mathbf{H}^{vT}$, $\mathbf{W}^v\mathbf{Y}^v_1$, $\mathbf{W}^v\mathbf{Y}^v_2$, and $\mathbf{W}^v(\mathbf{Q}^v)^{-1}\mathbf{Y}^v_3$. $\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT}$ and $\mathbf{X}^v\mathbf{H}^{vT}$ cost $O((n+m^v)d^{v2})$ and $O(m^vnd^v)$, respectively. Total cost of $\mathbf{W}^v\mathbf{Y}^v_1$, $\mathbf{W}^v\mathbf{Y}^v_2$ and $\mathbf{W}^v(\mathbf{Q}^v)^{-1}\mathbf{Y}^v_3$ is $O(m^vd^v + Kd^vn + d^{v2}n)$. K is the number of k nearest neighbors in graph. So, the cost for updating $\mathbf{W}^v$ is $O(Kd^vn + m^vnd^v)$. Normalize $\mathbf{W}^v$ and $\mathbf{H}^v$ in Equation (21) needs $O(nd^v + m^vd^v)$ computation. The main cost of updating $\mathbf{H}^v$ is on the calculation of $\mathbf{W}^{vT}\mathbf{X}^v$, $\mathbf{W}^{vT}\mathbf{W}^v\mathbf{H}^v$, $\mathbf{H}^v\mathbf{S}^v$ and $\mathbf{H}^v\mathbf{D}^v$. The cost on $\mathbf{W}^{vT}\mathbf{X}^v$ and $\mathbf{W}^{vT}\mathbf{W}^v\mathbf{H}^v$ are $O(m^vnd^v)$ and $O((n+m^v)d^{v2})$. While $\mathbf{H}^v\mathbf{S}^v$ and $\mathbf{H}^v\mathbf{D}^v$ need $O(Kd^vn)$ and $O(d^vn)$ computation

Table 1. Database Description

dataset	Size	Views	Features	Classes
Yale	165	3	2048/256/1024	15
ORL	400	3	2048/256/1024	40
FEI	700	3	2048/256/1024	50
3sources	169	3	3560/3631/3068	6
Texas	187	2	187/1703	5
Carotid	1378	2	17/17	2

respectively. The cost for updating $\mathbf{H}^v$ is $O(Kd^v n + m^v n d^v)$. Comparing with the cost of updating $\mathbf{W}^v$ and $\mathbf{H}^v$, the computation cost on updating $\mathbf{H}_c$ is ignorable. So, the final cost for FRSMNMF_N1 is $O(tVm^v nd^v)$. t is the number of iterations and V denotes the number of views.

For FRSMNMF_N2, the main difference of computation cost resides on the updating of $\mathbf{W}^v$. Except the computation cost on calculating $\mathbf{W}^v\mathbf{H}^v\mathbf{H}^{vT}$ and $\mathbf{X}^v\mathbf{H}^{vT}$, it also need to compute $\mathbf{Q}^v\mathbf{Y}_1^v$, $\mathbf{Q}^v\mathbf{Y}_2^v$ and $\mathbf{Y}_3^v$. The total computation cost on these additional terms is $O(Kd^v n + d^{v2} n)$. So, the cost for updating $\mathbf{W}^v$ in FRSMNMF_N2 is $O(Kd^v n + m^v n d^v)$. The final cost for FRSMNMF_N2 is $O(tVm^v nd^v)$.

5 EXPERIMENTS

5.1 Evaluation Metrics

Two widely used clustering metrics are adopted to evaluate the performance of the methods: **clustering accuracy (AC)** [49] and **normalized mutual information (NMI)** [40].

AC is defined as

$$\mathrm{AC} = \frac{\sum_{i=1}^{n} \delta(gnd_i, map(z_i))}{n}, \tag{34}$$

where n is the number of samples in each multiview dataset. gnd_i is the ground truth label of x_i. z_i corresponds to the label that is obtained by clustering methods due to the sample x_i. $map(\cdot)$ is an optimal mapping function [33] which can map the obtained label z_i to match the ground truth label provided by the datasets. $\delta(a, b)$ is an indicator function. $\delta(a, b) = 1$, when $a = b$, otherwise, $\delta(a, b) = 0$.

NMI is defined as follows:

$$\mathrm{NMI}(C, \hat{C}) = \frac{\mathrm{MI}(C, \hat{C})}{\sqrt{\mathrm{H}(C)\mathrm{H}(\hat{C})}}, \tag{35}$$

where C and $\hat{C}$ are two cluster sets, one for ground truth and another for the label obtained by clustering methods. H(C) and H($\hat{\mathrm{C}}$) denote the entropy of cluster sets C and $\hat{C}$. MI(C, $\hat{\mathrm{C}}$) is the mutual information between C and $\hat{C}$, which is defined as

$$\mathrm{MI}(C, \hat{C}) = \sum_{c_i \in C, \hat{c}_j \in \hat{C}} p(c_i, \hat{c}_j) \log_2 \frac{p(c_i, \hat{c}_j)}{p(c_i)p(\hat{c}_j)}, \tag{36}$$

where $p(c_i, c_j)$ denotes the joint probability that a randomly selected item belongs to c_i and c_j simultaneously, while $p(c_i)$ and $p(c_j)$ denote the probability that a randomly selected item belongs to c_i and c_j , respectively.

5.2 Datasets

In the experiments, six multiview datasets are used to evaluate the effectiveness of the proposed method on the clustering task. The statistics are summarized in Table 1.

(1) *Yale*[1] *Dataset.* This dataset contains 165 grayscale images collected from 15 people. Each person has 11 images with different facial expressions or configurations and all these images are normalized to 32 × 32 pixel array.
(2) *FEI Part 1*[2] *Dataset.* The FEI part 1 dataset is a subset of the original FEI database. This dataset contains 700 color images captured from 50 people. Each person has 14 images taken from different views. The original 640 × 480 resolution images were downsampled to 32 × 24 grayscale pixels.
(3) *ORL*[3] *Dataset.* This dataset has 400 grayscale face images collected from 40 people, each 10 images. These images are taken in different light conditions, with different facial expression and with/without glasses. From this database, two subsets are produced for testing.
(4) *Texas*[4] *Dataset.* The dataset contains 187 documents over the 5 labels (student, project, course, staff, faculty). The documents are described by 1,703 words in the content view, and by 187 links between them in cites views.
(5) *3sources* [16] *Dataset.* This dataset has 948 texts collected from three news sources, i.e., BBC, Reuters, and the Guardian. In this article, the experiments are conducted on the subset with 169 news articles and six topical labels that are reported in all three sources. The dimensions of this dataset's three views are 3,560; 3,631; and 3,068, respectively.
(6) *Carotid Dataset.* This dataset is an Electronic Medical Record dataset which has 1,378 records with 34 attributes. This dataset is a subset of the raw data collected from stroke screening and prevention project conducted by Shenzhen Second People's Hospital (Ethical approval Number: 20200116002). There are two classes in this dataset, i.e., abnormal carotid artery and normal carotid artery. Two views are constructed by splitting the whole attributes evenly, each has 17 attributes. The whole attributes are ranked by mean feature importance score of six feature importance scoring methods [19]. Odd-ranked attributes are used to construct one view and even-ranked attributes are used to construct the other view. The feature importance scores are provided by the work in Reference [19].

For **Yale**, **ORL,** and **FEI** datasets, the multiple views are the image intensity, Gabor [23], and LBP [37] with the dimensions of 1,024; 2,048; and 256, respectively.

The clustering results are obtained by conducting k-means clustering method on the learned representations of above algorithms. To avoid errors from randomness, all unsupervised methods were tested 10 times, and the average value is reported. To evaluate the performance of semi-supervised methods, for parameter analysis experiments, 10%, 20%, and 30% data points are labeled randomly for 5 times, and the average clustering results are reported. For comparison experiments, 10%, 20%, and 30% data points are labeled randomly for 20 times, and the average clustering results and statistically significant differences (p-value) are reported.

5.3 Comparative Algorithms

To evaluate the effectiveness of the proposed methods, they are compared with several representative NMF-based multi-view clustering approaches, including seven unsupervised ones and four semi-supervised ones. These methods are described as follows:

VAGNMF. Conduct GNMF [4] on each view individually and average feature of multiple views are treated as final representation.

[1] http://cvc.yale.edu/projects/yalefaces/yalefaces.html
[2] http://fei.edu.br/~cet/facedatabase.html
[3] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[4] http://lig-membres.imag.fr/grimal/data.html

VCGNMF. Conduct GNMF on each view individually and concatenate the features of multiple views as final representation.

LP-DiNMF [42]. The objective function of this method involves a diverse term which is defined to enforce the heterogeneity of the different views. Local geometric structure information is also utilized to improve the performance.

rNNMF [10]. This method attempts to deal with the noise and outliers among the views through defining a reconstruction term and a neighbor-structure-preserving term with $\ell_{2,1}$-norm.

MPMNMF_1 [46]. This method has adopted the Euclidean distance based pair-wise co-regularization to align multiple features. Graph regularization is constructed to encode the geometry information of each view.

MPMNMF_2 [46]. This method has adopted the kernel based pair-wise co-regularization to align multiple features. Graph regularization is constructed to encoding the geometry information of each view.

UDNMF [52]. This method factorizes the each view into three matrices and constrain the columns of product matrix of basis matrix and shared embedding matrix to be unit vector. Graph regularization is imposed on the coefficient matrix of each view. Centroid co-regularization is used to learn a common consensus matrix as the final representation.

MVCDMF [60]. This is a deep multi-view semi-NMF algorithm. "semi-nonnegative" means that it only constrains the shared coefficient matrix to be nonnegative. Graph regularization is also used to utilize the geometrical structure information of data. The shared coefficient matrix is used as the final representation.

MvDGNMF [26]. This is a deep multi-view NMF algorithm. It iteratively factorizes the coefficient matrix in each layer to get hierarchical representations of data. Similar to UDNMF, it adopts centroid co-regularization to fuse the multiple views to obtain the final representation.

AMVNMF [43]. This method factorizes each view in CNMF [30] framework and try to learn a consensus auxiliary matrix. The entries summation of each columns of auxiliary matrix is constrained to be 1. An auto-weighting strategy is imposed on the consensus learning terms.

MVCNMF [5]. This method factorizes each view in CNMF framework and align multiple views with Euclidean distance based pair-wise co-regularization. $\ell_{2,1}$-norm regularization is imposed on the auxiliary matrix in each view.

MVOCNMF [6]. This method factorizes each view in CNMF framework and align multiple views with Euclidean distance based pair-wise co-regularization. An orthonormality constraint is imposed on the auxiliary matrix in each view to normalize the scale of the feature.

GPSNMF [28]. This method is a semi-supervised partially shared MVNMF, which tries to make use of both distinct and shared information of multiple views. A graph regularization is also constructed for each view.

The results of above methods with optimal parameter settings (obtained by grid search) are reported. Note that, before conducting all methods, the data points of the original data are normalized as unit vectors. The source code of this paper is available at: https://github.com/GuoshengCui/FRSMNMF.

5.4 Convergence Analysis

To verify the convergence of FRSMNMF_N1 and FRSMNMF_N2, the convergence curves on six datasets with labeling ratios of 10%, 20%, and 30% are visualized in Figure 1. We also visualize the variation of clustering performance with different iterations. In each figure, the left y-axis denotes the objective function value, the right y-axis denotes the clustering accuracy and the x-axis is the iteration number. We can see that the proposed methods converge on all datasets and the clustering performances on all datasets become stable in 300 iterations.

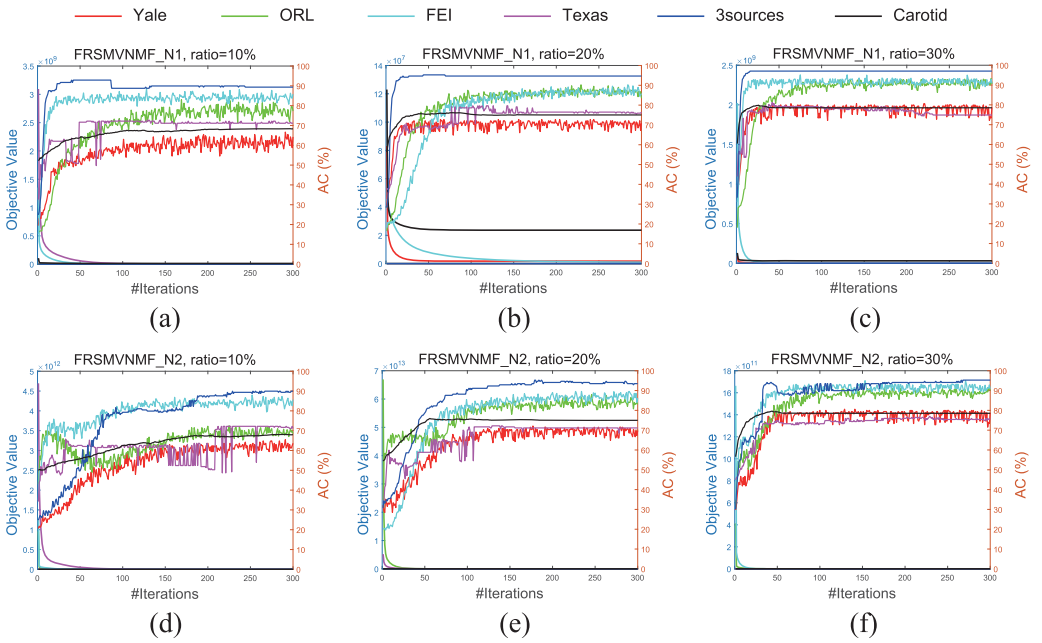


Fig. 1. Convergence curves and clustering performances of FRSMNMF_N1 and FRSMNMF_N2 with labeling ratios of 10%, 20% and 30% on six datasets. (Best viewed in color.)

5.5 Parameter Analysis and Comparisons

5.5.1 Parameter Analysis. There are three hyper-parameters, i.e., the graph regularization parameter α, the fusion regularization parameter β and k nearest neighbor parameter, in FRSMNMF_N1 and FRSMNMF_N2. The number of hyper-parameters in our methods is the same as the most unsupervised MVNMFs, such as LP-DiNMF, MPMNMF_1, MPMNMF_2 and UDNMF; and less than the most semi-supervised MVNMFs, such as MVCNMF, MVOCNMF, GPSNMF, and the method in [31]. Figure 2 demonstrates the influence of parameters α and β to FRSMNMF_N1 and FRSMNMF_N2 on six datasets with different labeling ratios, i.e., 10%, 20%, and 30%. From this figure, we can see that the performance varying trends of the proposed methods on α and β are very similar. This phenomenon indicates that the value of α/β is relatively stable on all datasets. The best parameter settings of the proposed methods on six datasets with different labeling ratios are demonstrated on Table 2. We can see that the value of α/β with best results are always locating in the sets of $\{1, 10, 100\}$ with α around 10^5 or 10^1.

Figure 3 has demonstrated the performance variation of GPSNMF, FRSMNMF_N1, and FRSMNMF_N2 versus the parameter k on six datasets with different labeling ratios. We can see that on most datasets, FRSMNMF_N1 and FRSMNMF_N2 have obtained better performance than GPSNMF on a large range of k. Basically, the varying trends of these three methods on six datasets are familiar.

5.5.2 Comparisons. In Table 3, the comparing results of our methods and several recently proposed semi-supervised MVNMFs are demonstrated. Statistical significant differences (p-value) between FRSMNMF_N1 and the other methods are also reported. AMVNMF, MVCNMF, and MVOCNMF are all CNMF-based MVNMFs, and their performances are obviously not as good as GPSNMF,

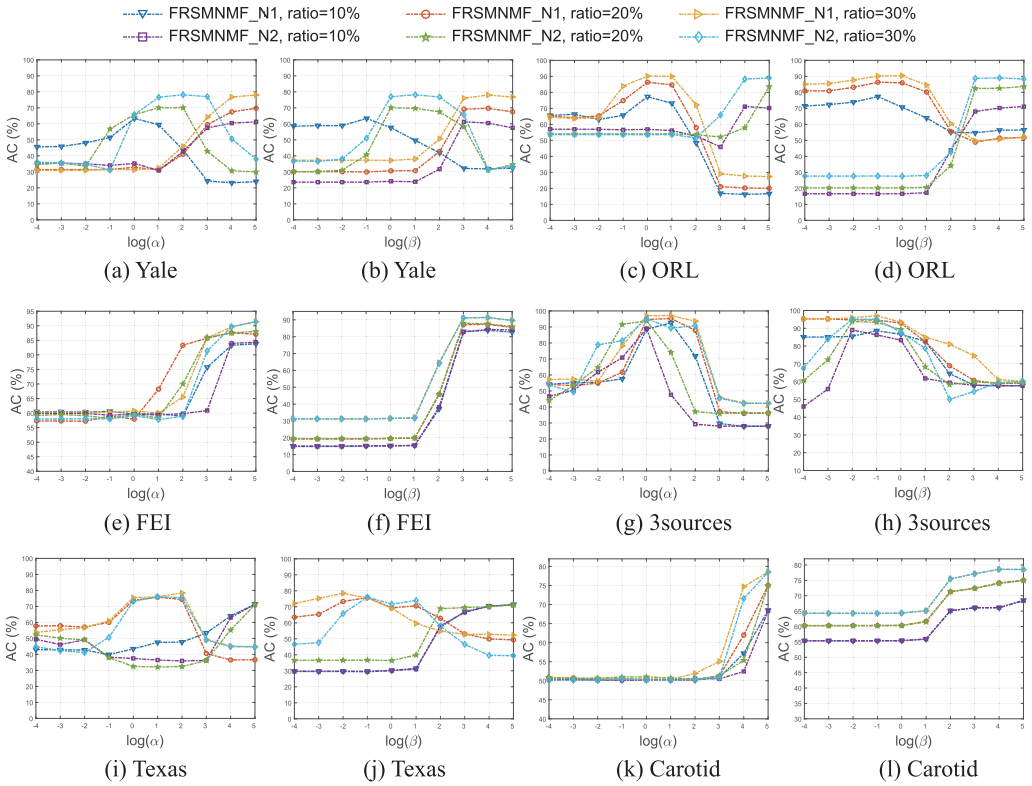


Fig. 2. Clustering performances of FRSMNMF_N1 and FRSMNMF_N2 versus parameters α and β with labeling ratios of 10%, 20%, and 30% on six datasets.

Table 2. The Best Parameter Settings of FRSMNMF_N1 and FRSMNMF_N2 on Six Datasets, the Labeling Ratios are 10%, 20%, and 30%

	FRSMNMF_N1			FRSMNMF_N2		
	10%	20%	30%	10%	20%	30%
Yale	(0, −1)	(5, 4)	(5, 4)	(2, 1)	(2, 0)	(2, 1)
ORL	(0, −1)	(0, −1)	(0, −1)	(3, 2)	(5, 5)	(5, 4)
FEI	(5, 4)	(5, 3)	(5, 4)	(5, 4)	(5, 3)	(5, 4)
3sources	(0, −1)	(0, −1)	(0, −1)	(0, −2)	(0, −2)	(0, −1)
Texas	(5, 5)	(1, −1)	(0, −1)	(5, 5)	(5, 5)	(1, −1)
Carotid	(5, 5)	(5, 5)	(5, 5)	(5, 5)	(5, 5)	(5, 5)

The values in brackets are (log(α), log(β)).

FRSMNMF_N1, and FRSMNMF_N2. One reason is that, the former three methods have not considered geometrical information of the data. The second reason is that, the former three methods have all failed to explore the discriminative information of the data, although they have considered intra-class compactness due to the basis of CNMF. On Yale, ORL, FEI, and 3sources datasets, the advantage of FRSMNMF_N1 is significant comparing to the other semi-supervised MVNMF methods

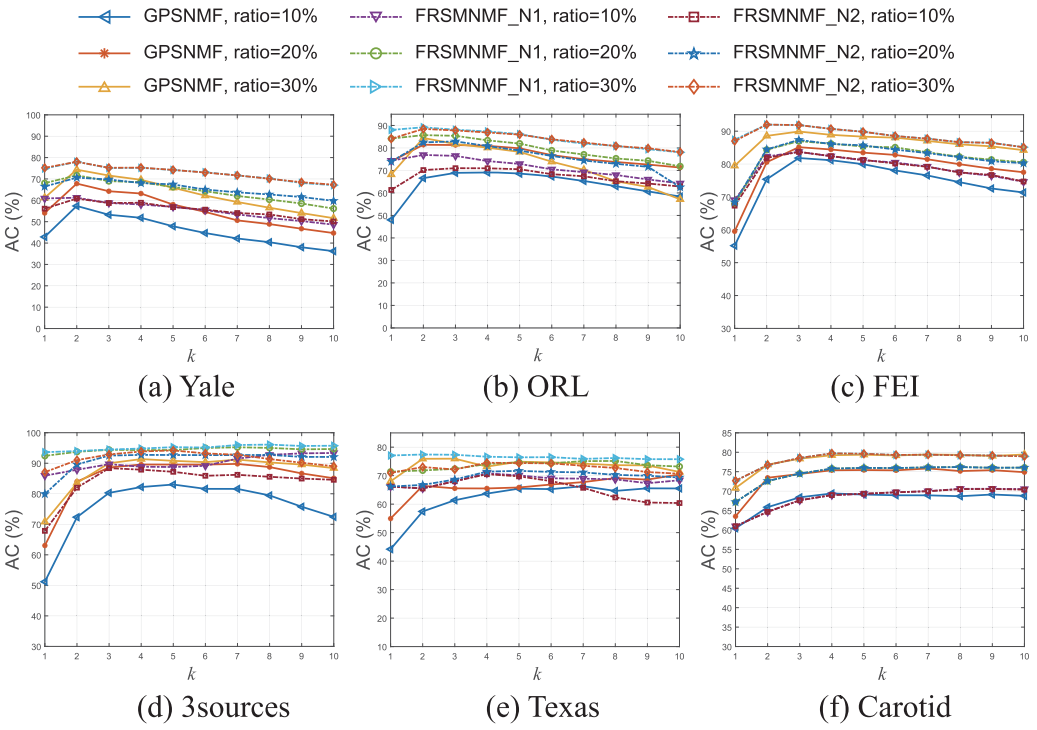


Fig. 3. Clustering performances of FRSMNMF_N1, FRSMNMF_N2, and GPSNMF versus parameter k nearest neighbors with labeling ratios of 10%, 20% and 30% on six datasets.

(except FRSMNMF_N2). On Texas and Carotid datasets, the advantage of FRSMNMF_N1 is significant comparing to AMVNMF, MVCNMF and MVOCNMF. On these two datasets, the performances of GPSNMF are approaching FRSMNMF_N1 as the labeling ratios increase. Overall, on all datasets, FRSMNMF_N1 usually report the best results. And in many cases, the performance differences of FRSMNMF_N1 and FRSMNMF_N2 are not that significant. Although FRSMNMF_N1 is slightly better than FRSMNMF_N2 which indicates that the former feature normalizing strategy works better.

Our methods are also compared with several recently proposed unsupervised MVNMFs. The comparing results are shown in Table 4. The last two rows are our methods with 10% labeled data points. Similarly, statistical significant differences (p-value) between FRSMNMF_N1 and the other methods are also reported. We can see that the advantage of the proposed FRSMNMF_N1 and FRSMNMF_N2 is very obvious comparing to the other methods. On ORL and 3sources datasets, FRSMNMF_N1 is superior to FRSMNMF_N2. On Yale, FEI, Texas and Carotid datasets, the performance differences of FRSMNMF_N1 and FRSMNMF_N2 are not that significant. And on all datasets, FRSMNMF_N1 is significantly better than the other unsupervised MVNMFs.

We also have tested the performance variation of all semi-supervised MVNMFs with different labeling ratios. The varying curves are demonstrated in Figure 4 with the mean results of all unsupervised methods as baseline. We can see that on the whole labeling ratios, GPSNMF, FRSMNMF_N1 and FRSMNMF_N2 are above the baseline. When the number of labeled data points is small, AMVNMF, MVCNMF and MVOCNMF sometimes behave worse than the baseline. Note that all unsupervised MVNMFs have considered geometrical information of the data. This means that

Table 3. Clustering Performance (AC(%) and NMI(%)) of Several Semi-supervised MVNMFs on Six Datasets with Different Labeling Ratios

Datasets	Yale						ORL					
	10%		20%		30%		10%		20%		30%	
	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value
AMVNMF [43]	49.97	<0.001	55.80	<0.001	62.95	<0.001	63.65	<0.001	68.02	<0.001	72.57	<0.001
MVCNMF [5]	50.93	<0.001	57.93	<0.001	63.56	<0.001	66.17	<0.001	70.07	<0.001	74.22	<0.001
MVOCNMF [6]	51.14	<0.001	58.03	<0.001	64.98	<0.001	66.40	<0.001	70.75	<0.001	75.30	<0.001
GPSNMF [28]	57.73	<0.001	67.50	<0.001	74.06	<0.001	68.82	<0.001	81.24	<0.001	84.09	<0.001
FRSMNMF_N2	**62.25**	0.777	*70.79*	0.484	*78.00*	0.903	*72.05*	<0.001	*82.65*	<0.001	*87.98*	0.015
FRSMNMF_N1	*62.01*	–	**71.41**	–	**78.07**	–	**76.40**	–	**85.48**	–	**89.24**	–
	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value
AMVNMF [43]	53.11	<0.001	59.62	<0.001	66.42	<0.001	79.27	<0.001	82.23	<0.001	84.81	<0.001
MVCNMF [5]	54.60	<0.001	61.62	<0.001	67.19	<0.001	81.07	<0.001	83.69	<0.001	86.01	<0.001
MVOCNMF [6]	54.59	<0.001	61.51	<0.001	68.13	<0.001	81.13	<0.001	84.02	<0.001	86.64	<0.001
GPSNMF [28]	59.27	<0.001	66.00	<0.001	72.05	<0.001	83.35	<0.001	*90.04*	<0.001	91.46	<0.001
FRSMNMF_N2	*61.11*	0.001	*68.03*	0.298	*75.15*	0.980	*83.97*	<0.001	89.77	<0.001	*92.33*	0.020
FRSMNMF_N1	**63.14**	–	**68.77**	–	**75.16**	–	**86.63**	–	**90.96**	–	**93.05**	–
Datasets	FEI						3sources					
	10%		20%		30%		10%		20%		30%	
	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value
AMVNMF [43]	56.65	<0.001	60.06	<0.001	66.36	<0.001	59.27	<0.001	65.83	<0.001	70.79	<0.001
MVCNMF [5]	59.19	<0.001	63.58	<0.001	69.60	<0.001	61.65	<0.001	69.83	<0.001	72.72	<0.001
MVOCNMF [6]	59.88	<0.001	63.47	<0.001	69.68	<0.001	57.44	<0.001	72.30	<0.001	81.12	<0.001
GPSNMF [28]	81.45	<0.001	84.56	<0.001	88.84	<0.001	80.79	<0.001	88.30	<0.001	91.05	<0.001
FRSMNMF_N2	*83.67*	0.506	**87.03**	0.428	*91.78*	0.796	*87.91*	0.020	*92.06*	0.083	*93.45*	<0.001
FRSMNMF_N1	**83.88**	–	*86.79*	–	**91.83**	–	**89.91**	–	**93.30**	–	**95.24**	–
	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value
AMVNMF [43]	76.13	<0.001	78.16	<0.001	82.36	<0.001	55.62	<0.001	61.02	<0.001	65.74	<0.001
MVCNMF [5]	77.43	<0.001	80.33	<0.001	84.35	<0.001	60.00	<0.001	66.76	<0.001	66.67	<0.001
MVOCNMF [6]	77.79	<0.001	80.16	<0.001	84.37	<0.001	51.77	<0.001	59.36	<0.001	68.13	<0.001
GPSNMF [28]	90.04	0.009	91.51	0.111	93.60	<0.001	67.39	<0.001	75.00	<0.001	78.56	<0.001
FRSMNMF_N2	*90.53*	0.739	**91.93**	0.659	*95.04*	0.895	*74.57*	<0.001	*82.21*	0.029	*84.04*	<0.001
FRSMNMF_N1	**90.88**	–	*91.84*	–	**95.06**	–	**79.39**	–	**84.70**	–	**88.06**	–
Datasets	Texas						Carotid					
	10%		20%		30%		10%		20%		30%	
	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value
AMVNMF [43]	45.55	<0.001	49.66	<0.001	55.73	<0.001	53.47	<0.001	56.61	<0.001	54.53	<0.001
MVCNMF [5]	64.61	<0.001	67.80	<0.001	70.77	<0.001	56.47	<0.001	60.49	<0.001	61.65	<0.001
MVOCNMF [6]	66.32	<0.001	67.91	<0.001	69.80	<0.001	56.43	<0.001	60.01	<0.001	65.45	<0.001
GPSNMF [28]	64.57	<0.001	70.12	<0.001	*75.25*	0.032	70.23	0.062	75.56	0.086	78.88	0.194
FRSMNMF_N2	*69.29*	0.224	*72.23*	0.002	73.54	<0.001	*71.04*	0.870	*75.91*	0.687	**79.19**	1.000
FRSMNMF_N1	**69.96**	–	**73.96**	–	**77.20**	–	**71.11**	–	**76.03**	–	**79.19**	–
	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value
AMVNMF [43]	20.31	<0.001	23.37	<0.001	28.57	<0.001	0.93	<0.001	3.41	<0.001	2.11	<0.001
MVCNMF [5]	32.27	<0.001	36.23	<0.001	41.71	<0.001	1.98	<0.001	4.88	<0.001	8.65	<0.001
MVOCNMF [6]	32.49	<0.001	36.14	<0.001	40.78	<0.001	1.90	<0.001	4.51	<0.001	7.06	<0.001
GPSNMF [28]	26.44	0.098	*39.96*	0.185	**53.80**	<0.001	12.50	0.125	20.29	0.333	26.00	0.501
FRSMNMF_N2	*28.87*	0.731	37.01	<0.001	*47.10*	0.525	*13.41*	0.945	*20.49*	0.656	**26.31**	1.000
FRSMNMF_N1	**29.16**	–	**41.52**	–	46.42	–	**13.45**	–	**20.70**	–	**26.31**	–

The best results are highlighted in **bold** and the second best results are in *italic*.

both label information and geometrical information of the data are important for the learning of meaningful representation.

5.5.3 Studying the Effects of Individual Views. To study the effect of each individual view, we apply k-means algorithm on the feature of each view and report the clustering results. To validate the clustering performance of FRSMNMF_N1 and FRSMNMF_N2, we select MVCNMF and MVOCNMF as baselines. Statistical significant differences (p-value) between FRSMNMF_N1 and the other methods are also reported. Table 5 provides the clustering results on Yale, ORL,

Table 4. Clustering Performance (AC(%) and NMI(%)) of the Proposed Methods and Several Unsupervised MVNMFs on Six Datasets

Datasets	Yale		ORL		FEI		3sources		Texas		Carotid	
	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value
VCGNMF [4]	54.24	<0.001	71.33	<0.001	69.54	<0.001	59.17	<0.001	62.51	<0.001	61.51	<0.001
VAGNMF [4]	50.00	<0.001	69.13	<0.001	68.83	<0.001	62.66	<0.001	54.28	<0.001	57.17	<0.001
LP-DiNMF [42]	50.85	<0.001	70.43	<0.001	69.54	<0.001	62.78	<0.001	55.35	<0.001	57.03	<0.001
rNNMF [10]	50.42	<0.001	65.58	<0.001	55.26	<0.001	62.07	<0.001	60.37	<0.001	57.89	<0.001
UDNMF [52]	47.88	<0.001	67.95	<0.001	74.71	<0.001	62.78	<0.001	54.71	<0.001	60.86	<0.001
MPMNMF_1 [46]	50.85	<0.001	70.40	<0.001	69.04	<0.001	72.13	<0.001	65.72	<0.001	57.20	<0.001
MPMNMF_2 [46]	50.12	<0.001	69.20	<0.001	68.76	<0.001	65.27	<0.001	61.71	<0.001	56.99	<0.001
MVCDMF [60]	53.52	<0.001	69.28	<0.001	68.84	<0.001	77.28	<0.001	56.74	<0.001	57.49	<0.001
MvDGNMF [26]	50.79	<0.001	70.45	<0.001	74.17	<0.001	71.66	<0.001	58.13	<0.001	56.79	<0.001
FRSMNMF_N2	**62.25**	0.777	*72.05*	<0.001	*83.67*	0.254	*87.91*	<0.001	*69.29*	0.050	*71.04*	0.333
FRSMNMF_N1	*62.01*	–	**76.40**	–	**83.88**	–	**89.91**	–	**69.96**	–	**71.11**	–
	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value
VCGNMF [4]	57.55	<0.001	*84.68*	<0.001	85.40	<0.001	58.68	<0.001	28.70	<0.001	3.97	<0.001
VAGNMF [4]	52.74	<0.001	83.67	<0.001	84.48	<0.001	56.50	<0.001	21.48	<0.001	2.36	<0.001
LP-DiNMF [42]	53.81	<0.001	84.65	<0.001	84.50	<0.001	56.37	<0.001	24.49	<0.001	2.34	<0.001
rNNMF [10]	52.56	<0.001	80.91	<0.001	74.13	<0.001	57.83	<0.001	27.89	<0.001	2.31	<0.001
UDNMF [52]	50.69	<0.001	82.22	<0.001	87.17	<0.001	58.41	<0.001	22.93	<0.001	4.44	<0.001
MPMNMF_1 [46]	54.39	<0.001	84.21	<0.001	85.09	<0.001	65.71	<0.001	**31.48**	<0.001	2.38	<0.001
MPMNMF_2 [46]	52.54	<0.001	84.46	<0.001	85.08	<0.001	58.61	<0.001	26.46	<0.001	2.59	<0.001
MVCDMF [60]	56.05	<0.001	84.27	<0.001	84.54	<0.001	71.59	<0.001	25.12	<0.001	2.16	<0.001
MvDGNMF [26]	53.44	<0.001	84.53	<0.001	87.15	<0.001	61.67	<0.001	28.39	<0.001	2.11	<0.001
FRSMNMF_N2	*61.11*	0.001	83.97	<0.001	*90.53*	0.372	*74.57*	<0.001	28.87	0.026	*13.41*	0.691
FRSMNMF_N1	**63.14**	–	**86.63**	–	**90.88**	–	**79.39**	–	*29.16*	–	**13.45**	–

The best results are highlighted in **bold** and the second best results are in *italic*.

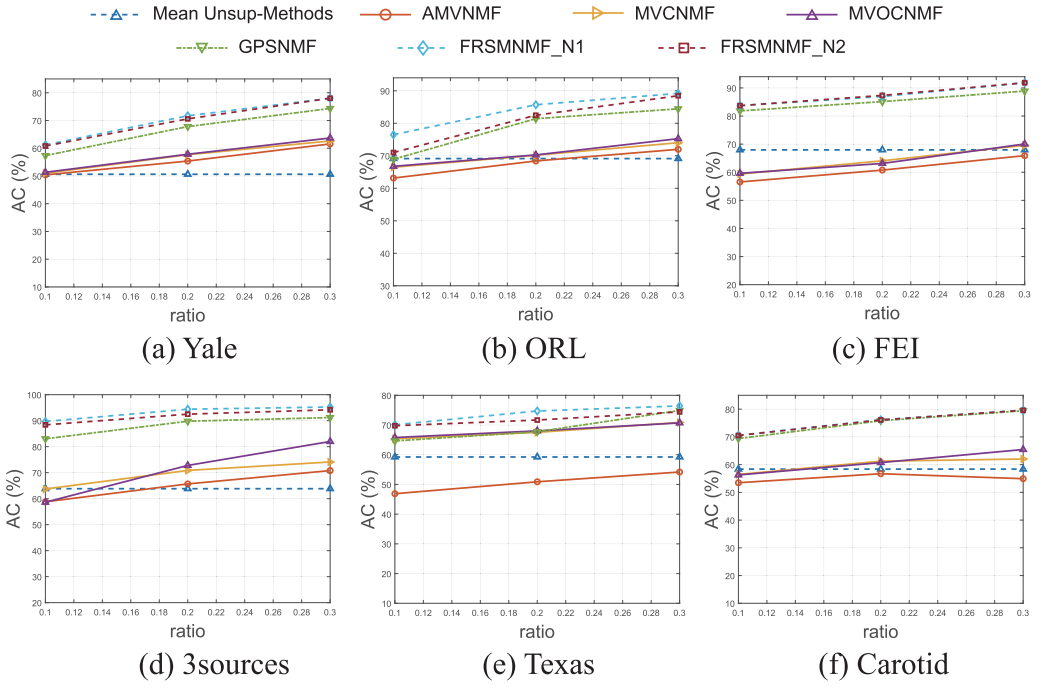


Fig. 4. Clustering performances of FRSMNMF_N1 and FRSMNMF_N2 versus different labeling ratios on six datasets.

Table 5. Individual View Clustering Performance (AC(%) and NMI(%)) of MVCNMF, MVOCNMF, FRSMNMF_N1, and FRSMNMF_N2 on Yale, ORL, 3sources and Carotid Datasets

Datasets	Yale						ORL					
	view1		view2		view3		view1		view2		view3	
	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value
MVCNMF [5]	60.95	<0.001	64.70	<0.001	56.56	<0.001	74.13	<0.001	74.54	<0.001	74.14	<0.001
MVOCNMF [6]	63.91	<0.001	63.97	<0.001	64.07	<0.001	75.06	<0.001	75.37	<0.001	74.42	<0.001
FRSMNMF_N2	77.84	0.954	**77.96**	0.954	**76.71**	0.954	88.01	0.133	88.41	0.133	87.35	0.133
FRSMNMF_N1	**77.98**	–	77.93	–	76.67	–	**89.16**	–	**89.61**	–	**88.17**	–
	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value
MVCNMF [5]	64.68	<0.001	68.03	<0.001	61.01	<0.001	85.99	<0.001	86.22	<0.001	86.02	<0.001
MVOCNMF [6]	67.64	<0.001	67.64	<0.001	67.63	<0.001	86.65	<0.001	86.70	<0.001	86.28	<0.001
FRSMNMF_N2	*74.69*	0.952	**74.93**	0.952	**73.88**	0.952	*92.24*	0.084	*92.55*	0.084	*91.77*	0.084
FRSMNMF_N1	**74.76**	–	*74.92*	–	*73.85*	–	**92.92**	–	**93.32**	–	**92.34**	–
Datasets	3sources						Carotid				–	
	view1		view2		view3		view1		view2		–	
	AC	p-value	AC	p-value	AC	p-value	AC	p-value	AC	p-value	–	–
MVCNMF [5]	71.65	<0.001	69.54	<0.001	68.89	<0.001	63.78	<0.001	64.73	<0.001	–	–
MVOCNMF [6]	81.75	<0.001	81.72	<0.001	81.74	<0.001	65.22	<0.001	65.22	<0.001	–	–
FRSMNMF_N2	*93.01*	0.004	*92.86*	0.004	*93.30*	0.004	**76.58**	1.000	**72.96**	1.000	–	–
FRSMNMF_N1	**95.18**	–	**94.58**	–	**94.87**	–	**76.58**	–	**72.96**	–	–	–
	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	NMI	p-value	–	–
MVCNMF [5]	67.38	<0.001	65.82	<0.001	64.77	<0.001	12.01	<0.001	10.71	<0.001	–	–
MVOCNMF [6]	69.13	<0.001	69.07	<0.001	69.09	<0.001	6.82	<0.001	6.82	<0.001	–	–
FRSMNMF_N2	*82.64*	<0.001	*82.54*	<0.001	*83.82*	<0.001	**21.81**	1.000	**15.96**	1.000	–	–
FRSMNMF_N1	**87.99**	–	**86.11**	–	**87.46**	–	**21.81**	–	**15.96**	–	–	–

The labeling ratios are all set to 30%. The best results are highlighted in **bold** and the second best results are in *italic*.

Table 6. Ablation Study of FRSMNMF_N1 and FRSMNMF_N2 on Yale, ORL, 3sources and Carotid Datasets

		Yale		ORL	
Norm	Method	AC	NMI	AC	NMI
None	FRSMNMF	61.55 ± 3.02	60.97 ± 2.17	69.69 ± 2.03	82.35 ± 1.40
$\|\|\mathbf{W}_{\cdot j}^{v}\|\|_2 = 1$	FRSMNMF_N1	*62.01 ± 2.31*	**63.14 ± 1.62**	**76.40 ± 1.39**	**86.63 ± 0.71**
	FRSMNMF_N1 w|o GR	58.43 ± 2.40	59.74 ± 1.91	71.24 ± 1.50	82.97 ± 0.73
	FRSMNMF_N1 w|o FR	49.78 ± 0.91	53.23 ± 0.54	67.97 ± 0.71	82.96 ± 0.29
$\|\|\mathbf{W}_{\cdot j}^{v}\|\|_1 = 1$	FRSMNMF_N2	**62.25 ± 3.01**	*61.11 ± 2.00*	*72.05 ± 1.51*	*83.97 ± 0.76*
	FRSMNMF_N2 w|o GR	44.48 ± 1.14	48.05 ± 0.71	62.56 ± 0.79	78.60 ± 0.40
	FRSMNMF_N2 w|o FR	51.47 ± 0.99	54.97 ± 0.86	68.64 ± 0.71	83.51 ± 0.28
		3sources		Carotid	
Norm	Method	AC	NMI	AC	NMI
None	FRSMNMF	84.35 ± 3.90	67.82 ± 5.68	69.07 ± 1.48	10.91 ± 1.74
$\|\|\mathbf{W}_{\cdot j}^{v}\|\|_2 = 1$	FRSMNMF_N1	**89.91 ± 2.49**	**79.39 ± 3.41**	**71.11 ± 1.49**	**13.45 ± 1.96**
	FRSMNMF_N1 w|o GR	85.04 ± 4.17	73.76 ± 3.63	61.11 ± 1.09	4.14 ± 0.58
	FRSMNMF_N1 w|o FR	60.55 ± 2.53	56.42 ± 1.86	55.76 ± 0.12	2.54 ± 0.07
$\|\|\mathbf{W}_{\cdot j}^{v}\|\|_1 = 1$	FRSMNMF_N2	*87.91 ± 2.71*	*74.57 ± 3.95*	*71.04 ± 1.46*	*13.41 ± 1.90*
	FRSMNMF_N2 w|o GR	53.68 ± 1.76	50.50 ± 1.60	58.48 ± 0.29	3.32 ± 0.17
	FRSMNMF_N2 w|o FR	59.10 ± 0.64	55.08 ± 0.85	57.17 ± 0.81	2.60 ± 0.41

The labeling ratio is set to 10%. The best results are highlighted in **bold** and the second best results are in *italic*.

3sources and Carotid datasets. The labeling ratios are all set to 30%. From Table 5, we can see that the proposed methods FRSMNMF_N1 and FRSMNMF_N2 outperform MVCNMF and MVOCNMF in each view. This phenomenon verifies the feature quality of each view in FRSMNMF_N1 and FRSMNMF_N2.

Table 7. Performances of FRSMNMF_N1 and FRSMNMF_N2 on Yale, ORL, FEI and Carotid datasets polluted by different levels of noises (10%, 20% and 30%). The labeling ratio is set to 20%.

Yale						
Noise	10%		20%		30%	
Method	AC	NMI	AC	NMI	AC	NMI
MVCNMF [5]	56.45 ± 1.69	60.18 ± 1.68	53.32 ± 1.71	57.85 ± 1.34	49.01 ± 1.46	53.40 ± 1.17
MVOCNMF [6]	56.61 ± 1.66	*60.44 ± 1.37*	53.82 ± 1.68	*58.22 ± 1.53*	50.02 ± 1.57	*54.22 ± 1.31*
FRSMNMF_N2	*57.59 ± 2.54*	56.16 ± 1.79	*54.60 ± 2.78*	52.14 ± 2.23	*52.37 ± 2.62*	50.03 ± 2.40
FRSMNMF_N1	**65.59 ± 1.86**	**64.85 ± 1.97**	**61.71 ± 1.94**	**61.15 ± 1.89**	**59.38 ± 2.55**	**58.39 ± 2.03**
ORL						
Noise	10%		20%		30%	
Method	AC	NMI	AC	NMI	AC	NMI
MVCNMF [5]	67.06 ± 1.11	81.10 ± 0.51	64.21 ± 1.09	78.64 ± 0.69	58.94 ± 1.16	74.74 ± 0.49
MVOCNMF [6]	66.92 ± 0.76	80.92 ± 0.45	64.12 ± 1.00	78.64 ± 0.53	58.90 ± 0.92	74.92 ± 0.41
FRSMNMF_N2	*78.42 ± 1.69*	*85.35 ± 1.12*	*75.70 ± 1.92*	*83.24 ± 1.12*	*71.82 ± 1.51*	*80.01 ± 1.14*
FRSMNMF_N1	**82.26 ± 1.45**	**88.68 ± 0.84**	**80.09 ± 1.20**	**86.91 ± 0.82**	**75.43 ± 1.24**	**83.69 ± 0.78**
FEI						
Noise	10%		20%		30%	
Method	AC	NMI	AC	NMI	AC	NMI
MVCNMF [5]	60.09 ± 0.85	77.25 ± 0.51	54.76 ± 1.20	73.23 ± 0.62	51.48 ± 0.87	70.29 ± 0.54
MVOCNMF [6]	59.85 ± 0.90	77.04 ± 0.60	54.63 ± 1.08	73.27 ± 0.51	51.08 ± 0.86	70.22 ± 0.57
FRSMNMF_N2	*83.68 ± 1.05*	*90.18 ± 0.70*	*79.54 ± 1.36*	*87.16 ± 0.76*	**76.93 ± 1.21**	**84.74 ± 0.75**
FRSMNMF_N1	**83.79 ± 1.13**	**90.21 ± 0.72**	**79.72 ± 1.59**	**87.20 ± 0.81**	*76.87 ± 1.11*	*84.70 ± 0.67*
Carotid						
Noise	10%		20%		30%	
Method	AC	NMI	AC	NMI	AC	NMI
MVCNMF [5]	59.88 ± 1.65	4.54 ± 1.06	59.07 ± 1.12	3.89 ± 0.78	58.62 ± 0.81	5.06 ± 0.65
MVOCNMF [6]	59.69 ± 1.54	4.39 ± 0.96	58.82 ± 1.12	3.71 ± 0.74	59.54 ± 0.56	4.18 ± 0.31
FRSMNMF_N2	**74.56 ± 1.02**	**18.28 ± 1.47**	*73.88 ± 1.36*	*17.26 ± 1.90*	*72.74 ± 1.97*	*15.70 ± 2.71*
FRSMNMF_N1	*74.54 ± 0.88*	*18.23 ± 1.28*	**73.95 ± 0.96**	**17.34 ± 1.39**	**73.23 ± 1.29**	**16.33 ± 1.86**

The best results are highlighted in **bold** and the second best results are in *italic*.

5.5.4 Ablation Study. In this subsection, we will further explore the influences of graph regularization, fusion regularization, the constraints $||\mathbf{W}^{v}_{\cdot j}||_2 = 1$ and $||\mathbf{W}^{v}_{\cdot j}||_1 = 1$. We denote FRSMNMF_N1 w|o GR as FRSMNMF_N1 without graph regularization, FRSMNMF_N1 w|o FR as FRSMNMF_N1 without fusion regularization, FRSMNMF_N2 w|o GR as FRSMNMF_N2 without graph regularization, FRSMNMF_N2 w|o FR as FRSMNMF_N2 without fusion regularization. We further denote FRSMNMF as FRSMNMF_N1 or FRSMNMF_N2 without the constraint $||\mathbf{W}^{v}_{\cdot j}||_2 = 1$ or $||\mathbf{W}^{v}_{\cdot j}||_1 = 1$. The performances of above methods on Yale, ORL, FEI and Carotid datasets with the labeling ratios of 10% are reported. The experimental results for the ablation study of FRSMNMF N1 and FRSMNMF N2 on Yale, ORL, 3sources and Carotid datasets are reported in Table 6. From Table 6, we can see that, on all datasets, the performances of FRSMNMF_N1 and FRSMNMF_N2 decrease without graph regularization or fusion regularization. This indicates that both graph regularization and fusion regularization contribute to the performances of FRSMNMF_N1 and FRSMNMF_N2. Comparing the results of FRSMNMF_N1, FRSMNMF_N2, and FRSMNMF, it can be seen that the performance of FRSMNMF decreases without the constraint $||\mathbf{W}^{v}_{\cdot j}||_2 = 1$ or $||\mathbf{W}^{v}_{\cdot j}||_1 = 1$. It means that the constraints $||\mathbf{W}^{v}_{\cdot j}||_2 = 1$ and $||\mathbf{W}^{v}_{\cdot j}||_1 = 1$ contribute to the performances of FRSMNMF_N1 and FRSMNMF_N2, respectively.

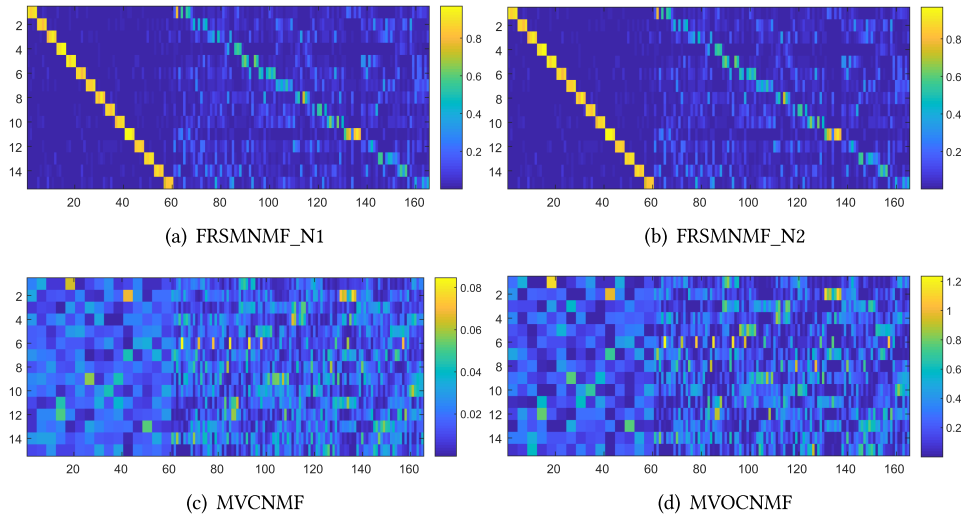
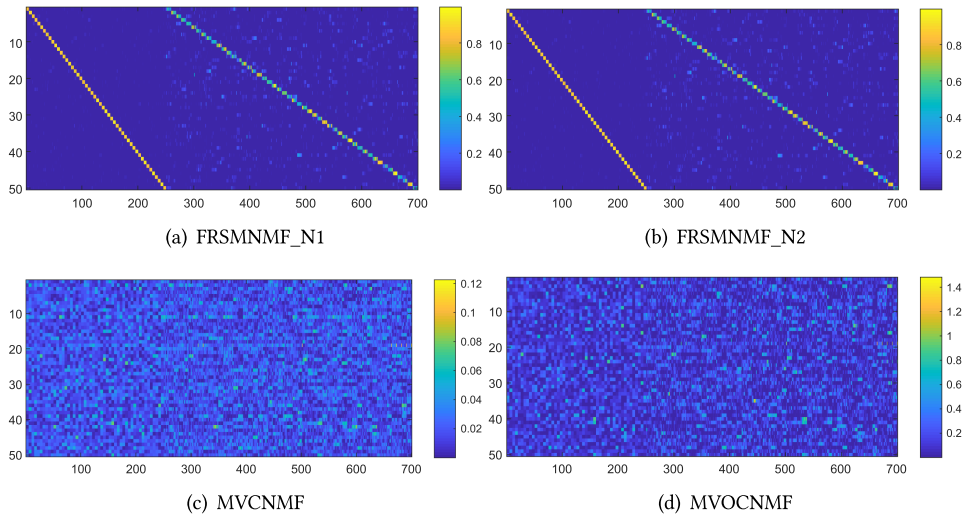


Fig. 5. The feature visualization of FRSMNMF_N1, FRSMNMF_N2, MVCNMF, and MVOCNMF on the Yale dataset. The labeling ratio is 30%. Rows are features and columns are data points.

5.5.5 Performance on Noise Datasets. In this section, we will evaluate the performances of the proposed methods on noise datasets. Specifically, we manually pollute Yale, ORL, FEI and Carotid datasets with different levels of noise. And the different levels of pollution are simulated by randomly discarding 10%, 20% , and 30% data points in each view. The performances of FRSMNMF_N1 and FRSMNMF_N2 are evaluated on above constructed noise datasets with MVCNMF and MVOCNMF as baselines. The labeling ratios of all methods are set to 20%. Before applying above methods, the missing data points in each view are filled with the average value of existing data points in the corresponding view. The experimental results on noise datasets are reported in Table 7. As can be seen from Table 7, the performances of all methods decrease with the increase of noise level. FRSMNMF_N1 is superior to FRSMNMF_N2 in most cases. On the Yale dataset, we find that the performance of FRSMNMF_N1 decreases dramatically, indicating that the constraint $||\mathbf{W}^v_{\cdot j}||_2 = 1$ is more robust-to-noise than the constraint $||\mathbf{W}^v_{\cdot j}||_1 = 1$.

5.5.6 Feature Visualization. In this section, we visualize the fused features learned by MVOCNMF, MVCNMF, FRSMNMF_N1, and FRSMNMF_N2 in Figure 5 (on the Yale dataset) and Figure 6 (on the FEI dataset). In this figure, rows are features and columns are data points. In Figures 5(a) and 5(b), the features of the first 60 data points, whose labels are given in the dataset, are with clear block diagonal structure. The block diagonal structure means that the distances between the data points with the same labels are minimized, approaching zero, while the distances between the data points with the different labels are maximized. In Figures 5(a) and 5(b), the feature matrix of the rest unlabeled data points also shows block diagonal structure, although with some noises. From Figures 5(c) and 5(d), we can see that, for the first 60 data points, although the distances of the data points with the same labels are zero (the data with same labels are represented with the same feature vectors), the distances of the data points with different classes are not minimized. This means that the label information is not effectively used in MVOCNMF and MVCNMF. The similar phenomenon can also be found in Figure 6, in which the labels of the top 250 data points are given.



(a) FRSMNMF_N1 (b) FRSMNMF_N2

(c) MVCNMF (d) MVOCNMF

Fig. 6. The feature visualization of FRSMNMF_N1, FRSMNMF_N2, MVCNMF, and MVOCNMF on FEI dataset. The labeling ratio is 30%. Rows are features and columns are data points.

6 CONCLUSION

In this article, a novel semi-supervised MVNMF framework with fusion regularization (FRSMNMF) is proposed. In our work, the discriminative term and the feature alignment term are fused as one regularizing term, this effectively enhances the learning of discriminative feature and reduces the number of hyper-parameters which makes the proposed framework more easy to fine tune than existing semi-supervised MVNMFs. The geometrical information is also considered by constructing graph regularizer for each view. To align multiple views effectively, two feature scale normalizing strategies are adopted, two corresponding specific implementations and iterative optimizing schemes of the proposed framework are presented. The effectiveness of our methods is evaluated by comparing with several state-of-the-art unsupervised and semi-supervised MVNMFs on six datasets.

ACKNOWLEDGMENTS

Thanks to doctor Lijie Ren of Shenzhen Second People's Hospital for helping to collect and organize the medical dataset used in this paper.

REFERENCES

[1] H. F. Bassani and A. F. R. Araujo. 2015. Dimension selective self-organizing maps with time-varying structure for subspace and projected clustering. *IEEE Transactions on Neural Networks and Learning Systems* 26, 3 (2015), 458–471.
[2] Steffen Bickel and Tobias Scheffer. 2004. Multi-view clustering. In *Proceedings of the 2004 SIAM International Conference on Data Mining*. SIAM, 19–26.
[3] Stephen Boyd and Lieven Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.
[4] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S. Huang. 2011. Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 8 (2011), 1548–1560.
[5] Hao Cai, Bo Liu, Yanshan Xiao, and LuYue Lin. 2019. Semi-supervised multi-view clustering based on constrained nonnegative matrix factorization. *Knowledge-Based Systems* 182 (2019), 104798.
[6] Hao Cai, Bo Liu, Yanshan Xiao, and LuYue Lin. 2020. Semi-supervised multi-view clustering based on orthonormality-constrained nonnegative matrix factorization. *Information Sciences* 536 (2020), 171–184.

[7] Xiao Cai, Feiping Nie, and Heng Huang. 2013. Multi-view k-means clustering on big data. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*. Citeseer.
[8] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. 2015. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 586–594.
[9] Guoqing Chao, Shiliang Sun, and Jinbo Bi. 2021. A survey on multi-view clustering. *IEEE Transactions on Artificial Intelligence* 2, 2 (2021), 146–168.
[10] Feiqiong Chen, Guopeng Li, Shuaihui Wang, and Zhisong Pan. 2019. Multiview clustering via robust neighboring constraint nonnegative matrix factorization. *Mathematical Problems in Engineering* 2019 (2019).
[11] Rui Chen, Yongqiang Tang, Wensheng Zhang, and Wenlong Feng. 2022. Deep multi-view semi-supervised clustering with sample pairwise constraints. *Neurocomputing* 500 (2022), 832–845.
[12] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. 2009. *Introduction to Algorithms*. MIT Press.
[13] Beilei Cui, Hong Yu, Tiantian Zhang, and Siwen Li. 2019. Self-weighted multi-view clustering with deep matrix factorization. In *Proceedings of the Asian Conference on Machine Learning*. 567–582.
[14] Guosheng Cui and Ye Li. 2022. Nonredundancy regularization based nonnegative matrix factorization with manifold learning for multiview data representation. *Information Fusion* 82 (2022), 86–98.
[15] Charles Elkan. 2003. Using the triangle inequality to accelerate k-means. In *Proceedings of the 20th International Conference on Machine Learning*, Vol. 3. 147–153.
[16] Derek Greene and Padraig Cunningham. 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd International Conference on Machine Learning*, Vol. 148. ACM, 377–384.
[17] Menglei Hu and Songcan Chen. 2018. Doubly aligned incomplete multi-view clustering. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*. 2262–2268.
[18] Ling Huang, Hong-Yang Chao, and Chang-Dong Wang. 2019. Multi-view intact space clustering. *Pattern Recognition* 86 (2019), 344–353.
[19] Xiaoxiang Huang, Guosheng Cui, Dan Wu, and Ye Li. 2020. A semi-supervised approach for early identifying the abnormal carotid arteries using a modified variational autoencoder. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, 595–600.
[20] Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv, and Xi Peng. 2020. Partially view-aligned clustering. *Advances in Neural Information Processing Systems* 33 (2020), 2892–2902.
[21] A. K. Jain, M. N. Murty, and P. J. Flynn. 1999. Data clustering: A review. *ACM Computing Surveys*. 31, 3 (1999).
[22] Yu Jiang, Jing Liu, Zechao Li, and Hanqing Lu. 2014. Semi-supervised unified latent factor learning with multi-view data. *Machine Vision and Applications* 25 (2014), 1635–1645.
[23] Martin Lades, Jan C. Vorbruggen, Joachim Buhmann, Jörg Lange, Christoph Von Der Malsburg, Rolf P. Wurtz, and Wolfgang Konen. 1993. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* 3 (1993), 300–311.
[24] Daniel D. Lee and H. Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755 (1999), 788–791.
[25] Daniel D. Lee and H. Sebastian Seung. 2000. Algorithms for non-negative matrix factorization. In *Proceedings of the Advances in Neural Information Processing Systems*. 556–562.
[26] Jianqiang Li, Guoxu Zhou, Yuning Qiu, Yanjiao Wang, Yu Zhang, and Shengli Xie. 2020. Deep graph regularized non-negative matrix factorization for multi-view clustering. *Neurocomputing* 390 (2020), 108–116.
[27] Naiyao Liang, Zuyuan Yang, Zhenni Li, Weijun Sun, and Shengli Xie. 2020. Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints. *Knowledge-Based Systems* 194 (2020), 105582.
[28] Naiyao Liang, Zuyuan Yang, Zhenni Li, Shengli Xie, and Chun-Yi Su. 2020. Semi-supervised multi-view clustering with graph-regularized partially shared non-negative matrix factorization. *Knowledge-Based Systems* 190 (2020), 105185.
[29] Naiyao Liang, Zuyuan Yang, Zhenni Li, Shengli Xie, and Weijun Sun. 2021. Semi-supervised multi-view learning by using label propagation based non-negative matrix factorization. *Knowledge-Based Systems* 228 (2021), 107244.
[30] Haifeng Liu, Zhaohui Wu, Xuelong Li, Deng Cai, and Thomas S. Huang. 2012. Constrained nonnegative matrix factorization for image representation. *Transactions on Pattern Analysis and Machine Intelligence* 34, 7 (2012), 1299–1311.
[31] Jing Liu, Yu Jiang, Zechao Li, Zhi-Hua Zhou, and Hanqing Lu. 2014. Partially shared latent factor learning with multiview data. *IEEE Transactions on Neural Networks and Learning Systems* 26, 6 (2014), 1233–1246.
[32] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. 2013. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 252–260.
[33] László Lovász and Michael D. Plummer. 2009. *Matching Theory*. Vol. 367. American Mathematical Society.
[34] Chun-Ta Lu, Lifang He, Hao Ding, Bokai Cao, and Philip S. Yu. 2018. Learning from multi-view multi-way data via structural factorization machines. In *Proceedings of the 2018 World Wide Web Conference*. 1593–1602.

[35] Peng Luo, Jinye Peng, Ziyu Guan, and Jianping Fan. 2018. Dual regularized multi-view non-negative matrix factorization for clustering. *Neurocomputing* 294 (2018), 1–11.
[36] Feiping Nie, Guohao Cai, Jing Li, and Xuelong Li. 2017. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Transactions on Image Processing* 27, 3 (2017), 1501–1511.
[37] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7 (2002), 971–987.
[38] Weihua Ou, Shujian Yu, Gai Li, Jian Lu, Kesheng Zhang, and Gang Xie. 2016. Multi-view non-negative matrix factorization by patch alignment framework with view consistency. *Neurocomputing* 204 (2016), 116–124.
[39] Nishant Rai, Sumit Negi, Santanu Chaudhury, and Om Deshmukh. 2016. Partial multi-view clustering using graph regularized NMF. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*. 2192–2197.
[40] Farial Shahnaz, Michael W. Berry, V. Paul Pauca, and Robert J. Plemmons. 2006. Document clustering using nonnegative matrix factorization. *Information Processing and Management* 42, 2 (2006), 373–386.
[41] Weixiang Shao, Lifang He, and S. Yu Philip. 2015. Multiple incomplete views clustering via weighted nonnegative matrix factorization with $L_{2,1}$ regularization. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 318–334.
[42] Jing Wang, Feng Tian, Hongchuan Yu, Chang Hong Liu, Kun Zhan, and Xiao Wang. 2017. Diverse non-negative matrix factorization for multiview data representation. *IEEE Transactions on Cybernetics* 48, 9 (2017), 2620–2632.
[43] Jing Wang, Xiao Wang, Feng Tian, Chang Hong Liu, Hongchuan Yu, and Yanbei Liu. 2016. Adaptive multi-view semi-supervised nonnegative matrix factorization. In *Proceedings of the International Conference on Neural Information Processing*. Springer, 435–444.
[44] Senhong Wang, Jiangzhong Cao, Fangyuan Lei, Qingyun Dai, Shangsong Liang, and Bingo Wing-Kuen Ling. 2021. Semi-supervised multi-view clustering with weighted anchor graph embedding. *Computational Intelligence and Neuroscience* 2021 (2021).
[45] Xiaobo Wang, Xiaojie Guo, Zhen Lei, Changqing Zhang, and Stan Z. Li. 2017. Exclusivity-consistency regularized multi-view subspace clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 923–931.
[46] Xiumei Wang, Tianzhen Zhang, and Xinbo Gao. 2019. Multiview clustering based on non-negative matrix factorization and pairwise measurements. *IEEE Transactions on Cybernetics* 49, 9 (2019), 3333–3346.
[47] Chang Xu, Dacheng Tao, and Chao Xu. 2013. A survey on multi-view learning. *CoRR* abs/1304.5634 (2013).
[48] Jinglin Xu, Junwei Han, and Feiping Nie. 2016. Discriminatively embedded k-means for multi-view clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5356–5364.
[49] Wei Xu, Xin Liu, and Yihong Gong. 2003. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 267–273.
[50] Xiaojun Yang, Weizhong Yu, Rong Wang, Guohao Zhang, and Feiping Nie. 2020. Fast spectral clustering learning with hierarchical bipartite graph for large-scale data. *Pattern Recognition Letters* 130 (2020), 345–352.
[51] Yan Yang and Hao Wang. 2018. Multi-view clustering: A survey. *Big Data Mining and Analytics* 1, 2 (2018), 83–107.
[52] Zuyuan Yang, Naiyao Liang, Wei Yan, Zhenni Li, and Shengli Xie. 2020. Uniform distribution non-negative matrix factorization for multiview clustering. *IEEE Transactions on Cybernetics* 51, 6 (2020), 3249–3262.
[53] Zuyuan Yang, Yong Xiang, Kan Xie, and Yue Lai. 2016. Adaptive method for nonsmooth nonnegative matrix factorization. *IEEE Transactions on Neural Networks and Learning Systems* 28, 4 (2016), 948–960.
[54] Zuyuan Yang, Huimin Zhang, Naiyao Liang, Zhenni Li, and Weijun Sun. 2023. Semi-supervised multi-view clustering by label relaxation based non-negative matrix factorization. *The Visual Computer* 39, 4 (2023), 1409–1422.
[55] Shan Zeng, Xiuying Wang, Hui Cui, Chaojie Zheng, and David Feng. 2017. A unified collaborative multikernel fuzzy clustering for multiview data. *IEEE Transactions on Fuzzy Systems* 26, 3 (2017), 1671–1687.
[56] Hongyuan Zha, Xiaofeng He, Chris Ding, Ming Gu, and Horst D. Simon. 2002. Spectral relaxation for k-means clustering. In *Proceedings of the Advances in Neural Information Processing Systems*. 1057–1064.
[57] Changqing Zhang, Qinghua Hu, Huazhu Fu, Pengfei Zhu, and Xiaochun Cao. 2017. Latent multi-view subspace clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4279–4287.
[58] DongPing Zhang, YiHao Luo, YuYuan Yu, QiBin Zhao, and GuoXu Zhou. 2022. Semi-supervised multi-view clustering with dual hypergraph regularized partially shared non-negative matrix factorization. *Science China Technological Sciences* 65, 6 (2022), 1349–1365.
[59] Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. 2018. Binary multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 7 (2018), 1774–1782.
[60] Handong Zhao, Zhengming Ding, and Yun Fu. 2017. Multi-view clustering via deep matrix factorization. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*.

[61] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. 2017. Multi-view learning overview: Recent progress and new challenges. *Information Fusion* 38 (2017), 43–54.
[62] Wei Zhao, Cai Xu, Ziyu Guan, and Ying Liu. 2020. Multiview concept learning via deep matrix factorization. *IEEE Transactions on Neural Networks and Learning Systems* 32, 2 (2020), 814–825.
[63] X. Zhu, S. Zhang, Y. Li, J. Zhang, L. Yang, and Y. Fang. 2019. Low-rank sparse subspace for spectral clustering. *IEEE Transactions on Knowledge and Data Engineering* 31, 8 (2019), 1532–1543.
[64] Linlin Zong, Xianchao Zhang, Long Zhao, Hong Yu, and Qianli Zhao. 2017. Multi-view clustering via multi-manifold regularized non-negative matrix factorization. *Neural Networks* 88 (2017), 74–89.

Received 19 October 2022; revised 11 September 2023; accepted 8 March 2024