Comprehensive Multiview Representation Learning via Deep Autoencoder-Like Nonnegative Matrix Factorization

Haonan Huang, *Graduate Student Member, IEEE*, Guoxu Zhou, *Member, IEEE*, Qibin Zhao, *Senior Member, IEEE*, Lifang He, *Member, IEEE*, and Shengli Xie, *Fellow, IEEE*

***Abstract*— Learning a comprehensive representation from multiview data is crucial in many real-world applications. Multiview representation learning (MRL) based on nonnegative matrix factorization (NMF) has been widely adopted by projecting high-dimensional space into a lower order dimensional space with great interpretability. However, most prior NMF-based MRL techniques are shallow models that ignore hierarchical information. Although deep matrix factorization (DMF)-based methods have been proposed recently, most of them only focus on the consistency of multiple views and have cumbersome clustering steps. To address the above issues, in this article, we propose a novel model termed deep autoencoder-like NMF for MRL (DANMF-MRL), which obtains the representation matrix through the deep encoding stage and decodes it back to the original data. In this way, through a DANMF-based framework, we can simultaneously consider the multiview consistency and complementarity, allowing for a more comprehensive representation. We further propose a one-step DANMF-MRL, which learns the latent representation and final clustering labels matrix in a unified framework. In this approach, the two steps can negotiate with each other to fully exploit the latent clustering structure, avoid previous tedious clustering steps, and achieve optimal clustering performance. Furthermore, two efficient iterative optimization algorithms are developed to solve the proposed models both with theoretical convergence analysis. Extensive experiments on five benchmark datasets demonstrate the superiority of our approaches against other state-of-the-art MRL methods.**

Manuscript received 30 August 2022; revised 8 April 2023 and 22 June 2023; accepted 4 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62073087, Grant 62071132, Grant 62006045, Grant 61973090, and Grant U1911401; in part by the National Science Foundation under Grant MRI 2215789 and Grant IIS 1909879; in part by Lehigh University under Grant S00010293 and Grant 001250; and in part by the China Scholarship Council (CSC) under Grant 202208440315. *(Corresponding author: Guoxu Zhou.)*

Haonan Huang is with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China, and also with the RIKEN Center for Advanced Intelligence Project (AIP), Tokyo 103-0027, Japan (e-mail: libertyhhn@foxmail.com).

Guoxu Zhou and Shengli Xie are with the School of Automation and the Key Laboratory of Intelligent Detection and The Internet of Things in Manufacturing, Ministry of Education, Guangdong University of Technology, Guangzhou 510006, China (e-mail: gx.zhou@gdut.edu.cn; shlxie@gdut.edu.cn).

Qibin Zhao is with the RIKEN Center for Advanced Intelligence Project (AIP), Tokyo 103-0027, Japan, and also with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China (e-mail: qibin.zhao@riken.jp).

Lifang He is with the Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA 18015 USA (e-mail: lih319@lehigh.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2023.3304626.

Digital Object Identifier 10.1109/TNNLS.2023.3304626

***Index Terms*— Clustering, deep matrix factorization (DMF), multiview learning, nonnegative matrix factorization (NMF).**

NOMENCLATURE

V	Number of views.
m	Number of layers.
p_i	ith layer size.
n	Number of samples.
K	Number of clusters.
α^v	Weight factor of vth view.
$\mathbf{X}^v \in \mathbb{R}^{l_v \times n}$	Data matrix of the vth view.
$\mathbf{W}_i^v$	Basis matrix in the ith layer of the vth view.
$\mathbf{H}_m^v \in \mathbb{R}^{p_v \times n}$	mth layer representation matrix of the vth view.
$\mathbf{H}^* \in \mathbb{R}^{p_v \times n}$	Consensus representation matrix.
$\mathbf{L}^v \in \mathbb{R}^{n \times n}$	Graph Laplacian matrix of the vth view.
$\mathbf{F}^v \in \mathbb{R}^{K \times K}$	Centroid matrix of the vth view.
$\mathbf{Y} \in \mathbb{R}^{K \times n}$	Clustering label matrix.

I. INTRODUCTION

SINCE information of data typically comes from many sources, a growing amount of data are created from multiple views in the actual situations in recent years. For instance, an image usually contains different characteristics, including color, texture, and edge; a video consists of both visuals and audio, and a piece of content may be narrated in a variety of languages [1], [2]. All of these are referred to as multiview data, where each view represents a learning challenge but also has its own biases. The occurrence of multiview data in a natural and frequent manner spawned a new learning paradigm known as multiview representation learning (MRL) [3]. The challenge of learning representations (or features) of multiview data that assist in extracting useful information while creating prediction models is addressed by MRL. It has gained popularity in the field of unsupervised learning and has been used in numerous practical applications, including object recognition [4] and medical image analysis [5], among others. A basic approach of MRL is to concatenate all views into one single view and use the single-view learning method. The disadvantages of this strategy are that overfitting occurs on relatively small datasets and that the individual statistical

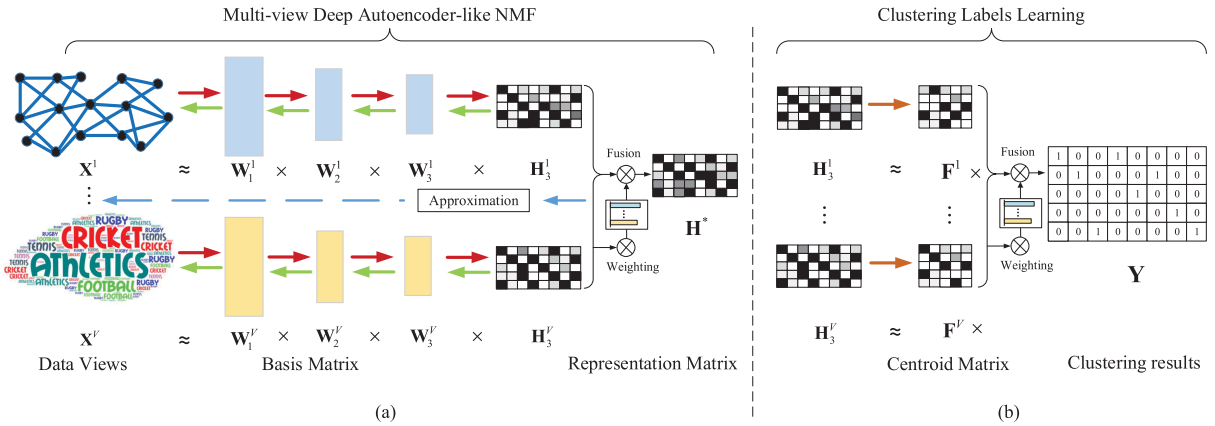


Fig. 1. (a) Overview of the proposed DANMF-MRL model. We take the data of two views as an example, and there are two key components, including the deep autoencoder-like NMF and the learning of common representation $\mathbf{H}^*$. The first part is composed of the deep encoder component (shown as the flow with red arrows) and the deep decoder component (shown as the flow with green arrows). In the second part, we can learn the common representation through the fusion strategy. (b) Learning of the consensus final clustering results $\mathbf{Y}$ is accomplished by the multiplication of V view-specific centroid matrices $\mathbf{F}$.

feature of each view is neglected [6]. In the last several years, many strategies (such as kernel methods [7], [8], subspace clustering [9], [10], spectral learning [11], and autoencoder and matrix factorization [12], [13]) have been employed to tackle the MRL problem.

Among these approaches, one of the most widely used MRL methods is based on nonnegative matrix factorization (NMF) [14]. The sparse and nonnegative constraints in matrix factorization make it possible to extract more discriminative features, which can eventually result in semantically meaningful representations of objects that are based on their component parts. Inspired by this, Liu et al. [15] developed a joint NMF-based multiview method with consensus term to learn the underlying clustering space embedded in multiple views. Liang et al. [16] designed a novel NMF method with co-orthogonal constraints to capture the multiview diversity attributes. Shi et al. [17] unified spectral clustering and NMF into a novel model to improve computational efficiency. Despite the impressive performance of NMF-based MRL methods, we observe that most of them suffer from the following drawbacks: 1) they are most shallow methods, which cannot discover the complex hierarchical and structural multiview information; 2) they cannot well simultaneously capture the multiview consistency and complementary information, resulting in inferior performance; and 3) most methods generally regard representation learning and clustering as two separated steps, which means that the learned representation may not be appropriate for the clustering task.

In addition, autoencoder is a popular unsupervised representation learning algorithm that is composed of two parts: an encoder and a decoder. The former maps input data into a latent space using latent representations, while the latter reconstructs input data using latent representations [18], [19], [20], [21]. However, these methods fail to consider the nonnegative structure of data, which leads to inadequate interpretability of learned representations. Moreover, the neural network-based techniques lack theoretical analysis for convergence in an unsupervised setting.

To address the aforementioned limitations, in this article, we propose a novel framework termed deep autoencoder-like NMF for MRL (DANMF-MRL). Different from existing deep autoencoder-based works, we first leverage the deep autoencoder-like NMF (DANMF) structure to discover the hierarchical mappings between the original data and the final representation to improve the model's learning ability. Specifically, as shown in Fig. 1(a), our method uses a unified framework to perform multiview consistency learning (with the deep encoder component) and each view information decoding (with the deep decoder component), allowing us to balance the consistency and complementarity among multiple views and learn a common intact representation. We further improve DANMF-MRL and develop another variant, termed one-step DANMF-MRL (OS-DANMF-MRL), to perform multiview matrix factorization with clustering partition generation in a unified model, as shown in Fig. 1(b). To the best of our knowledge, DANMF-MRL is the first effort to benefit MRL in a DANMF way. As a result, DANMF might provide some fresh insights to the MRL community. The contributions and novelties of ours can be summed up as follows when compared to well-studied matrix factorization-based MRL methods.

1) Aiming at MRL, we propose a new MRL model, i.e., DANMF-MRL, which considers loss terms quantified by the deep encoder and deep decoder parts, to consider the consistency and complementarity among multiple views simultaneously and obtain a final comprehensive representation. The new model also includes a graph regularization, which can retain the original data's manifold topology to improve its ability of representation learning.
2) We further develop the OS-DANMF-MRL, which unifies the latent representation learning and clustering assignments in a one-step model. This joint optimization technique successfully exploits multiview representation and clustering structure information, resulting in more precise clustering results. Two iterative algorithms are proposed to solve the resulting optimization problems with mathematical convergence analysis. Extensive experiments on five multiview benchmark datasets demonstrate the effectiveness of our models compared to other state-of-the-art techniques.

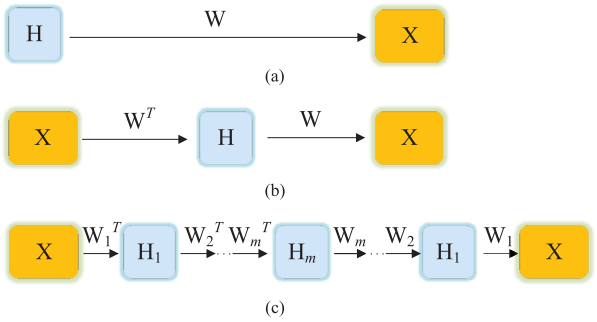


Fig. 2. Factorization process of different NMF-based methods: (a) traditional shallow NMF [29], (b) NSED presented in [23], and (c) DANMF [26].

The rest of this article is organized as follows. Section II outlines the background and related works of MRL. The proposed DANMF-MRL, OS-DANMF-MRL, and two corresponding algorithms with convergence analysis are introduced in Section III. Section IV shows the experiment results with evaluation. Section V concludes this article. Some results in this article were presented in [22].[1]

II. Background and Related Works

In this section, we first revisit the single-view autoencoder-like NMF models and then review the previous matrix factorization-based MRL methods that are related to our works. The Nomenclature lists the main notations used in this article.

A. Single-View Autoencoder-Like NMF

Given a single-view data matrix $\mathbf{X} \in \mathbb{R}^{l\times n}$ with l dimension and n samples, as shown in Fig. 2(a), NMF decomposes it into two nonnegative subspace matrices $\mathbf{W} \in \mathbb{R}^{l\times p}$ and $\mathbf{H} \in \mathbb{R}^{p\times n}$, where p is the dimension of subspace. Its cost function can be expressed as

$$\min_{\mathbf{W},\mathbf{H}} \mathcal{O}_{\text{NMF}} = \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 \quad \text{s.t. } \mathbf{W}, \mathbf{H} \geq 0. \tag{1}$$

To improve the representation ability of traditional NMF, Sun et al. [23] integrated a decoder component and an encoder component into a unified objective function, as shown in Fig. 2(b), called nonnegative symmetric encoder–decoder (NSED), as follows:

$$\min_{\mathbf{W},\mathbf{H}} \mathcal{O}_{\text{NSED}} = \underbrace{\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2}_{\text{decoder}} + \underbrace{\left\|\mathbf{H} - \mathbf{W}^T\mathbf{X}\right\|_F^2}_{\text{encoder}}$$
$$\text{s.t. } \mathbf{W}, \mathbf{H} \geq 0. \tag{2}$$

The symmetry of the encoder and decoder imposes a natural soft orthogonality restriction on the basis matrix $\mathbf{W}$. The corresponding representation matrix has superior clustering performance when the column vectors of the learned basis matrix are more independent in the NMF frame [24]. However, there is only one layer mapping between the original data and the low-dimensional space in NSED, which cannot capture the complex hierarchical patterns of real-world data. Inspired by the development of deep learning [25], Ye et al. [26] extended NSED to a multilayer version and designed the DANMF to automatically uncover the hierarchical mappings of attributes to facilitate clustering tasks. Mathematically, the data matrix $\mathbf{X} \in \mathbb{R}^{l\times n}$ is factorized into m nonnegative basis matrices $\{\mathbf{W}_1, \ldots, \mathbf{W}_m\}$ and a nonnegative representation matrix $\mathbf{H}_m$. The process of multilayer factorization can be expressed as

$$\begin{aligned} \mathbf{X} &\approx \mathbf{W}\mathbf{H} \\ \mathbf{X} &\approx \mathbf{W}_1\mathbf{W}_2\mathbf{H}_2 \\ &\vdots \\ \mathbf{X} &\approx \mathbf{W}_1, \ldots, \mathbf{W}_m\mathbf{H}_m. \end{aligned} \tag{3}$$

As shown in Fig. 2(c), the deep decoder stage obtains the best basis matrices and low-dimensional representation matrix $\mathbf{H}_m$ to reconstruct the input data $\mathbf{X}$. The deep encoder stage fine-tunes submatrices to transform the original data $\mathbf{X}$ into the distributed representation $\mathbf{H}_m$. While DANMF has the NSED algorithm capabilities mentioned above, this architecture enables it to better inherit the learning ability of deep autoencoder. In addition to community detection, autoencoder-like NMF has previously proven to be effective in a variety of additional applications, including remote sensing [27] and data clustering [28]. Nevertheless, the advanced autoencoder-like NMF-based methods fail to solve the problem of MRL.

B. Related Works

In recent years, numerous multiview representation methods have been proposed in order to tackle pervasive multiview data analysis problems. K-nearest neighbor (KNN) is a classical learning method to predict the labels of test samples and has been successful in multiple applications, including graph learning [30], spectral learning, and multiview learning [31], [32]. A novel joint model called NMF-KNN is designed to tackle the problem of dataset imbalance [33]. In [34], an effective KNN variant model is proposed to handle the issue of the problematic selection of k by automatically learning the optimal k values for different test data. Recently, an adaptive KNN and multiview spectral learning unified framework have been presented to fully exploit the heterogeneous information among different views [35] Canonical correlation analysis (CCA)-based models are one of the existing methods for MRL. Andrew et al. [36] introduced deep CCA to learn complex nonlinear transformations of two data views to produce highly linearly correlated representations. CCA-based methods are all seeking to multiview correlation, and they are limited to only two-view data.

Multiview subspace clustering methods are to find a low-rank self-representation of multiview data that represent its segmentation in a low-dimensional space [37], [38]. To discover the common underlying information shared by different views, Zhang et al. [39] proposed a novel method called latent multiview subspace clustering. Recently, motivated by the tensor learning [40], Xiao et al. [41] proposed a prior knowledge regularized multiview self-representation to make use of labeled information and high-order correlation to learn an accurate multiview self-representation tensor. Based on the characteristics of the Laplacian graph

[1]Upon our short conference version [22], this article designs a novel OS-DANMF-MRL method (see Section III-C), analyzes the convergence of the proposed algorithms (see Section III-F), and conducts more comprehensive experiments to validate the advantages of key components (see Section IV).

and its eigenvectors, spectral learning methods determine a division of the affinity graph. Nie et al. [11] presented a parameter-free spectral learning framework for multiview clustering and semisupervised classification. In order to learn the common affinity matrix as the clustering outcome, Zhu et al. [42] presented a unified one-step spectral clustering model. Wang et al. [43] introduced a general graph-based multiview clustering with a novel multiview automatically fusion technique. However, spectral-based methods are relatively limited in practical applications due to their high complexity, especially for large-scale datasets. Combining deep CCA with autoencoder, Wang et al. [18] designed a deep canonically correlated autoencoders (DCCAE) to find the multiview correlation-based representation and achieve head-to-head comparison. To learn disentangled visual multiview representations, Xu et al. [19] designed a novel generative model for multiview clustering called multi-VAE. More recently, Zhang et al. [20] proposed an unsupervised multiview learning framework, which consists of inner and outer autoencoder networks to learn a compact representation. Yang et al. [21] presented a robust multiview clustering, which adopts a unified autoencoder structure to face the partially view-unaligned and sample missing cases. However, existing deep autoencoder methods suffer from the following drawbacks: 1) they neglect the nonnegativity nature of data and unavoidably hamper interpretability of the learned representations and 2) despite being rooted in neural network, these algorithms lack theoretical convergence guarantees in unsupervised scenarios.

NMF [29], a powerful tool that can learn low-dimensional representations with more discriminative features, has also drawn a lot of interest in MRL and has shown significant promise for clustering. Liu et al. [15] constructed coefficient matrices by executing NMF on each data view and then drove them toward a consensus representation. Based on this, a uniform distribution multiview NMF (MultiNMF) was proposed to reduce distribution divergences among different views by jointly learning a common representation [44]. Orthogonality is extensively investigated in single-view NMF, with promising performance [24]. To exploit the complementary information between different views, a diversity-induced NMF (DiNMF) with orthogonal constraints is presented in [45]. Following this, Liang et al. [16] designed an NMF with co-orthogonal constraints (NMFCC) on both the basis matrix and representation matrix to enhance the clustering performance.

However, the abovementioned methods are shallow frameworks that rely on a linear mapping from the original data matrix to the newly mapped space [46], [47]. In real-world applications, this linear mapping is unable to mine the hierarchical structure that might express several degrees of abstraction to represent objects [48], [49]. To solve this problem, Zhao et al. [50] proposed a deep matrix factorization multiview framework [DMF-based multiview clustering (DMVC)] to maximize the consistency among multiple views. Based on this work, Huang et al. [51] presented a novel deep MF model to automatically learn the weights of different views. In [52], a semisupervised partially shared deep matrix factorization (PSDMF) is presented to extract the correlated and uncorrelated features of multiview data. Zhao et al. [53] also developed a deep multiview concept learning technique in a semisupervised paradigm. Nevertheless, as they depend on the labels, the suggested objective functions cannot guarantee their success. Recently, Chang et al. [54] proposed a deep concept factorization for MRL to catch the comprehensive and hierarchical information of the multiview data. Luong et al. [55] designed a deep NMF model with both orthonormality and graph constraints to learn complementary and compatible information embedded in multiview data.

Despite the good representation performance of the aforementioned deep matrix factorization-based MRL approaches, they mostly only consider the consistency of multiview data and neglect the complementary information existing in each view. In addition, they regard performing representation learning and clustering operations as two separate steps. Because the resulting representations are fixed after unsupervised extraction, they cannot be refined further to provide better clustering performance. To cope with these issues, the proposed methods minimize DANMF and multiview fusion terms, which help us to explore the multiview complementary and consistent information and achieve one-step clustering label learning. After learning the multiview representation matrices, the proposed OS-DANMF-MRL adopts clustering partition level fusion, avoiding the addition k-means or spectral clustering step in previous works [50], [54]. Since the following clustering phase and the representation learning step are connected, our method can achieve better clustering performance. Section III will introduce two proposed models and their corresponding optimizations.

III. Proposed Methods

A. DANMF-MRL Model

In the multiview setting, let us denote $\mathbf{X} = \{\mathbf{X}^1, \ldots, \mathbf{X}^V\}$ as the input data, and we can naturally extend the DANMF as follows:

$$\begin{aligned}\min_{\mathbf{W}_1^v,\ldots,\mathbf{W}_m^v,\mathbf{H}_m^v} \mathcal{O} = \sum_{v=1}^{V} \Big(&\left\|\mathbf{X}^v - \mathbf{W}_1^v\mathbf{W}_2^v,\ldots,\mathbf{W}_m^v\mathbf{H}_m^v\right\|_F^2 \\ &+ \left\|\mathbf{H}_m^v - \mathbf{W}_m^{v\,T},\ldots,\mathbf{W}_2^{vT}\mathbf{W}_1^{vT}\mathbf{X}^v\right\|_F^2 \\ &+ \lambda \mathrm{tr}\left(\mathbf{H}_m^v\mathbf{L}^v\mathbf{H}_m^{v\,T}\right)\Big) \\ \text{s.t. } \mathbf{W}_1^v,\ldots,\mathbf{W}_m^v,\mathbf{H}_m^v \geq 0 \end{aligned} \tag{4}$$

where $\mathbf{L}^v$ denotes the graph Laplacian matrix $\mathbf{L}^v = \mathbf{D}^v - \mathbf{A}^v$ and $\mathbf{D}^v$ is a diagonal matrix whose entries are row sums of $\mathbf{A}^v$, i.e., $\mathbf{D}_{jj}^v = \sum_q \mathbf{A}_{jq}^v$. According to [56], 0–1 weighting is one of the most prevalent methods for defining the weight matrix $\mathbf{A}^v$ on the graph, which is defined as

$$\mathbf{A}_{jq}^v = \begin{cases} 1, & \text{if } \boldsymbol{x}_q^v \in \mathcal{L}\left(\boldsymbol{x}_j^v\right) \text{ or } \boldsymbol{x}_j^v \in \mathcal{L}\left(\boldsymbol{x}_q^v\right) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\mathcal{L}(\boldsymbol{x}_q^v)$ is the set of k-nearest neighbor (k-NN) of the example $\boldsymbol{x}_q^v$.

According to [3], MRL assumes that different patterns among distinct viewpoints should be compatible. Following that, we try to find each viewpoint's relevance while forcing

their latent representation to be a common representation. Let $\mathbf{H}^*$ denote the common representation of all the views, and the multiview fusion term can be formulated as

$$\min_{\mathbf{H}^*,\alpha^v} \sum_{v=1}^{V} (\alpha^v)^\gamma \left\|\mathbf{H}_m^v - \mathbf{H}^*\right\|_F^2 \text{ s.t. } \mathbf{H}^* \geq 0, \quad \sum_{v=1}^{V} \alpha^v = 1 \tag{6}$$

where α^v is the weight for the vth view and γ is the hyperparameter to control the weights distribution. For (6), we can find that we will get equal weight factors when $\gamma \to \infty$. Also, when $\gamma \to 1^+$, we will give 1 to the weight factor of the view with least $\|\mathbf{H}_m^v - \mathbf{H}^*\|_F^2$ value and 0 to the weights of the other views. Employing such an approach, we can avoid the trivial solution to the weight distribution of the different views when $\gamma > 1$. In addition, in the light of [15], we need to normalize factors to make the disagreement measure appropriate for different $\mathbf{H}_m^v$ values when compared to the same consensus $\mathbf{H}^*$ and improve the accuracy of the approximation. Let $\mathbf{\Psi}_i^v = \mathbf{W}_1^v\mathbf{W}_2^v, \ldots, \mathbf{W}_i^v$, and the normalization imposed on $\mathbf{H}_m^v$ and $\mathbf{\Psi}_i^v$ is achieved by

$$\mathbf{H}_m^v \leftarrow \mathbf{H}_m^v(\mathbf{Q}^v)^{-1}, \quad \mathbf{\Psi}_m^v \leftarrow \mathbf{\Psi}_m^v\mathbf{Q}^v \tag{7}$$

where $\mathbf{Q}^v = \text{diag}(\sum_i(\mathbf{H}_m^v)_{i,1}, \sum_i(\mathbf{H}_m^v)_{i,2}, \ldots, \sum_i(\mathbf{H}_m^v)_{i,k})$.

Therefore, considering the above two aspects, the proposed DANMF-MRL can be finally formulated as

$$\begin{aligned}\min_{\substack{\mathbf{W}_1^v,\ldots,\mathbf{W}_m^v,\\ \mathbf{H}_m^v,\mathbf{H}^*,\alpha^v}} \mathcal{O} = \sum_{v=1}^{V} \Big(&\left\|\mathbf{X}^v - \mathbf{W}_1^v\mathbf{W}_2^v,\ldots,\mathbf{W}_m^v\mathbf{H}_m^v\right\|_F^2 \\ &+ \left\|\mathbf{H}_m^v - \mathbf{W}_m^{v\,T},\ldots,\mathbf{W}_2^{vT}\mathbf{W}_1^{vT}\mathbf{X}^v\right\|_F^2 \\ &+ \lambda\text{tr}\left(\mathbf{H}_m^v\mathbf{L}^v\mathbf{H}_m^{v\,T}\right) + (\alpha^v)^\gamma\left\|\mathbf{H}_m^v - \mathbf{H}^*\right\|_F^2\Big) \\ \text{s.t. } \mathbf{W}_1^v,\ldots,\mathbf{W}_m^v,\mathbf{H}_m^v,\mathbf{H}^* \geq 0, \quad &\sum_{v=1}^{V}\alpha^v = 1.\end{aligned} \tag{8}$$

For our objective function, the first term is the deep decoder stage, which obtains the best basis matrices and low-dimensional representation matrix $\mathbf{H}_m^v$ to reconstruct the input data $\mathbf{X}^v$. The second term is the deep encoder stage, which fine-tunes submatrices to transform the original data $\mathbf{X}^v$ into the distributed representation $\mathbf{H}_m^v$. With the DANMF, we also achieve latent orthogonality constraint on basis matrices, which would be helpful in discovering the more discriminative representation. The third term respects the intrinsic geometrical structure of original data. Also, the fourth term fuses multiple $\mathbf{H}_m^v$ into a consensus $\mathbf{H}^*$ with updated weights. In general, the deep encoder and the fusion components learn a consistent representation together. At the same time, the representation can be decoded back to the original data space by the deep decoder, and the manifold structure of each view is preserved by graph regularization, which together ensures the view-specific information of the data, i.e., the complementarity among multiple views. To summarize, the proposed model can automatically balance the consistency and complementarity between different views and learn a comprehensive representation through encoding the original data and decoding back.

B. DANMF-MRL Optimization

In this section, we mainly discuss how to solve (8). For brevity, we first define $\mathbf{P}^v = \mathbf{H}_m^v - \mathbf{H}^*$ and the objective function of DANMF-MRL can be further rewritten with the properties of matrix trace as follows:

$$\begin{aligned}\mathcal{O} = \sum_{v=1}^{V}\Big\{\text{tr}\Big(&\mathbf{X}^{v\,T}\mathbf{X}^v + \mathbf{H}_m^{v\,T}\mathbf{H}_m^v - 4\mathbf{X}^{v\,T}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{H}_m^v \\ &+ \mathbf{H}_m^{v\,T}\mathbf{\Phi}_{i+1}^{v\,T}\mathbf{W}_i^{vT}\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{H}_m^v \\ &+ \mathbf{X}^{vT}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{\Phi}_{i+1}^{v\,T}\mathbf{W}_i^{vT}\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{X}^v\Big) \\ &+ \lambda\text{tr}\left(\mathbf{H}_m^v\mathbf{L}^v\mathbf{H}_m^{v\,T}\right) + (\alpha^v)^\gamma\text{tr}(\mathbf{P}^{v\,T}\mathbf{P}^v)\Big\} \\ \text{s.t. } \mathbf{W}_i^v, \mathbf{H}_m^v, \mathbf{H}^* \geq 0, \quad &\sum_{v=1}^{V}\alpha^v = 1 \ \forall i = 1,\ldots,m\end{aligned} \tag{9}$$

where $\mathbf{\Psi}_{i-1}^v = \mathbf{W}_1^v\mathbf{W}_2^v, \ldots, \mathbf{W}_{i-1}^v$ and $\mathbf{\Phi}_{i+1}^v = \mathbf{W}_{i+1}^v, \ldots, \mathbf{W}_m^v$. It is worth noting that we denote $\mathbf{\Psi}_0^v = \mathbf{I}$ and $\mathbf{\Phi}_{m+1}^v = \mathbf{I}$.

1) Update Rule for Hidden Matrix $\mathbf{W}_i^v (m \geq i \geq 0)$: We minimize $\mathcal{O}$ over $\mathbf{W}_i^v$, with $\mathbf{H}_m^v$, $\mathbf{H}^*$, and α^v being fixed. Then, the objective function in (9) can be reduced to

$$\begin{aligned}\mathcal{O}(\mathbf{W}_i^v) = \text{tr}\Big(&-4\mathbf{X}^{v\,T}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{H}_m^v \\ &+ \mathbf{H}_m^{v\,T}\mathbf{\Phi}_{i+1}^{v\,T}\mathbf{W}_i^{vT}\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{H}_m^v \\ &+ \mathbf{X}^{vT}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{\Phi}_{i+1}^{v\,T}\mathbf{W}_i^{vT}\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{X}^v\Big) \\ \text{s.t. } &\mathbf{W}_i^v \geq 0.\end{aligned} \tag{10}$$

Let $\mathbf{\Theta}_i^v$ be the Lagrange multiplier for nonnegative constraint on $\mathbf{W}_i^v$, resulting in the following Lagrangian function:

$$\begin{aligned}&\mathcal{L}(\mathbf{W}_i^v, \mathbf{\Theta}_i^v) \\ &= \text{tr}\Big(-4\mathbf{X}^{v\,T}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{H}_m^v \\ &\quad + \mathbf{H}_m^{v\,T}\mathbf{\Phi}_{i+1}^{v\,T}\mathbf{W}_i^{vT}\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{H}_m^v \\ &\quad + \mathbf{X}^{vT}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{\Phi}_{i+1}^{v\,T}\mathbf{W}_i^{vT}\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{X}^v - \mathbf{\Theta}_i^v\mathbf{W}_i^{v\,T}\Big).\end{aligned} \tag{11}$$

By setting the partial derivative of $\mathcal{L}(\mathbf{W}_i^v, \mathbf{\Theta}_i^v)$ with respect to $\mathbf{W}_i^v$ to $\mathbf{0}$, we have

$$\mathbf{\Theta}_i^v = -4\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{X}^v\mathbf{H}_m^{v\,T}\mathbf{\Phi}_{i+1}^{v\,T} + 2\mathbf{\Pi}_i^v \tag{12}$$

where

$$\begin{aligned}\mathbf{\Pi}_i^v = &\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{H}_m^v\mathbf{H}_m^{v\,T}\mathbf{\Phi}_{i+1}^{v\,T} \\ &+ \mathbf{\Psi}_{i-1}^{v\,T}\mathbf{X}^v\mathbf{X}^{v\,T}\mathbf{\Psi}_{i-1}^v\mathbf{W}_i^v\mathbf{\Phi}_{i+1}^v\mathbf{\Phi}_{i+1}^{v\,T}.\end{aligned} \tag{13}$$

From the complementary slackness condition of the Karush–Kuhn–Tucker (KKT) conditions, we obtain

$$\mathbf{\Theta}_i^v\mathbf{W}_i^v = \left(-4\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{X}^v\mathbf{H}_m^{v\,T}\mathbf{\Phi}_{i+1}^{v\,T} + 2\mathbf{\Pi}_i^v\right)\mathbf{W}_i^v = \mathbf{0}. \tag{14}$$

Equation (14) is the fixed-point equation that the solution must satisfy at convergence. By solving this equation, we derive the following updating rule for $\mathbf{W}_i^v$:

$$\mathbf{W}_i^v = \mathbf{W}_i^v \odot \frac{2\mathbf{\Psi}_{i-1}^{v\,T}\mathbf{X}^v\mathbf{H}_m^{v\,T}\mathbf{\Phi}_{i+1}^{v\,T}}{\mathbf{\Pi}_i^v} \tag{15}$$

where $\odot$ denotes the elementwise product.

2) Update Rule for Representation Matrix $\mathbf{H}_m^v$: The objective function in (9) can be reduced to (16) by fixing all the variables except for $\mathbf{H}_m^v$

$$\begin{aligned}\mathcal{O}(\mathbf{H}_m^v) = \operatorname{tr}\Big(&\mathbf{H}_m^{v\,T}\mathbf{H}_m^v + \mathbf{H}_m^{v\,T}\boldsymbol{\Psi}_m^{v\,T}\boldsymbol{\Psi}_m^v\mathbf{H}_m^v \\ &- 4\mathbf{X}^{vT}\boldsymbol{\Psi}_m^v\mathbf{H}_m^v + \lambda\mathbf{H}_m^v\mathbf{L}^v\mathbf{H}_m^{v\,T} \\ &+ (\alpha^v)^\gamma\left(\mathbf{H}_m^v - \mathbf{H}^*\right)^T\left(\mathbf{H}_m^v - \mathbf{H}^*\right)\Big) \\ \text{s.t. } &\mathbf{H}_m^v \geq 0. \end{aligned} \tag{16}$$

We set the graph Laplacian matrix $\mathbf{L}^v = \mathbf{D}^v - \mathbf{S}^v$ and follow the similar derivation process of the update rule for $\mathbf{W}_i$, and the update rule for $\mathbf{H}_m^v$ is obtained as follows:

$$\mathbf{H}_m^v = \mathbf{H}_m^v \odot \frac{2\boldsymbol{\Psi}_m^{v\,T}\mathbf{X}^v + \lambda\mathbf{H}_m^v\mathbf{S}^v + (\alpha^v)^\gamma\mathbf{H}^*}{\boldsymbol{\Psi}_m^{v\,T}\boldsymbol{\Psi}_m^v\mathbf{H}_m^v + \lambda\mathbf{H}_m^v\mathbf{D}^v + (1+(\alpha^v)^\gamma)\mathbf{H}_m^v}. \tag{17}$$

3) Update Rule for Consensus Matrix $\mathbf{H}^*$: Optimizing the consensus matrix $\mathbf{H}^*$ amounts to solving the following function:

$$\mathcal{O}(\mathbf{H}^*) = \sum_{v=1}^{V}(\alpha^v)^\gamma \operatorname{tr}\Big(\left(\mathbf{H}_m^v - \mathbf{H}^*\right)^T\left(\mathbf{H}_m^v - \mathbf{H}^*\right)\Big), \quad \text{s.t. } \mathbf{H}^* \geq 0. \tag{18}$$

We take the derivative of (18) with respect to $\mathbf{H}^*$, set it to 0, and obtain an exact solution for $\mathbf{H}^*$

$$\mathbf{H}^* = \frac{\sum_v^V (\alpha^v)^\gamma \mathbf{H}_m^v}{\sum_v^V (\alpha^v)^\gamma}. \tag{19}$$

4) Update Rule for Weight Factors α^v: To optimize the weight factor, we only consider the term that is relevant to α^v. Equation (9) can be rewritten as

$$\mathcal{O}(\alpha^v) = \sum_{v=1}^{V}(\alpha^v)^\gamma \mathcal{P}^v, \quad \text{s.t. } \sum_{v=1}^{V}\alpha^v = 1 \tag{20}$$

where $\mathcal{P}^v = \|\mathbf{P}^v\|_F^2$. Therefore, let Λ^v be the multiplier, and the Lagrange function of (20) is

$$\mathcal{L}(\alpha^v, \Lambda^v) = \sum_{v=1}^{V}(\alpha^v)^\gamma \mathcal{P}^v - \Lambda^v\left(\sum_{v=1}^{V}\alpha^v - 1\right). \tag{21}$$

Then, we set the derivative of (21) with respect to α^v to zero and have

$$\alpha^v = \left(\frac{\Lambda^v}{\gamma\mathcal{P}^v}\right)^{\frac{1}{\gamma-1}}. \tag{22}$$

After substituting the resultant α^v in (22) into the constraint $\sum_{v=1}^V \alpha^v = 1$, we can obtain the optimal solution as follows:

$$\alpha^v = \frac{(\mathcal{P}^v)^{\frac{1}{1-\gamma}}}{\sum_{v=1}^V (\mathcal{P}^v)^{\frac{1}{1-\gamma}}}. \tag{23}$$

C. One-Step DANMF-MRL

Although the proposed DANMF-MRL in Section III-A demonstrates promising clustering performance in MRL, we find that it works by learning a consensus representation matrix $\mathbf{H}^*$ and then applying k-means on this matrix to generate the final clustering labels. This suggests that these two steps lack negotiation to reach optimality. To address the above concern, we propose an OS-DANMF-MRL, which directly learns the discrete clustering labels. The k-means method attempts to split the representation matrix $\mathbf{H}^*$ into K distinct clusters, each defined by its centroid, which can be expressed as

$$\begin{aligned}&\min_{\mathbf{F},\mathbf{Y}} \|\mathbf{H}^* - \mathbf{F}\mathbf{Y}\|_F^2 \\ &\text{s.t. } \mathbf{Y}_{kj} \in \{0,1\}, \quad \sum_{k=1}^{K}\mathbf{Y}_{kj} = 1 \ \forall j = 1,\ldots,n \end{aligned} \tag{24}$$

where $\mathbf{F} \in \mathbb{R}^{K\times K}$ is a centroid matrix and $\mathbf{Y} \in \mathbb{R}^{K\times n}$ is an indicator matrix, which stores the final clustering results about samples. Thus, we design a one-step multiview fusion term by combining (6) and (24) as follows:

$$\begin{aligned}&\min_{\mathbf{F}^v,\mathbf{Y},\alpha^v} \sum_{v=1}^{V}(\alpha^v)^\gamma \left\|\mathbf{H}_m^v - \mathbf{F}^v\mathbf{Y}\right\|_F^2 \\ &\text{s.t. } \sum_{v=1}^{V}\alpha^v = 1, \quad \mathbf{Y}_{kj} \in \{0,1\}, \quad \sum_{k=1}^{K}\mathbf{Y}_{kj} = 1 \ \forall j = 1,\ldots,n. \end{aligned} \tag{25}$$

Because the hard partition is unique in multiview clustering tasks, we employ $\mathbf{F}^v$ to represent the view-specific centroid matrix and a consensus label indicator matrix $\mathbf{Y}$ across all views. Combining this one-step fusion strategy with the objective function of DANMF (4) gives rise to our OS-DANMF-MRL, which minimizes the objective function as follows:

$$\begin{aligned}\min_{\substack{\mathbf{W}_1^v,\ldots,\mathbf{W}_m^v,\\ \mathbf{H}_m^v,\mathbf{F}^v,\mathbf{Y},\alpha^v}} \mathcal{O} = \sum_{v=1}^{V}\Big(&\left\|\mathbf{X}^v - \mathbf{W}_1^v\mathbf{W}_2^v,\ldots,\mathbf{W}_m^v\mathbf{H}_m^v\right\|_F^2 \\ &+ \left\|\mathbf{H}_m^v - \left(\mathbf{W}_m^v\right)^T,\ldots,\left(\mathbf{W}_2^v\right)^T\left(\mathbf{W}_1^v\right)^T\mathbf{X}^v\right\|_F^2 \\ &+ \lambda\operatorname{tr}\left(\mathbf{H}_m^v\mathbf{L}^v\left(\mathbf{H}_m^v\right)^T\right) + (\alpha^v)^\gamma\left\|\mathbf{H}_m^v - \mathbf{F}^v\mathbf{Y}\right\|_F^2\Big) \\ \text{s.t. } \mathbf{W}_i^v, \mathbf{H}_m^v \geq 0, \quad &\sum_{v=1}^{V}\alpha^v = 1 \ \forall i = 1,\ldots,m \\ \mathbf{Y}_{kj} \in \{0,1\}, \quad &\sum_{k=1}^{K}\mathbf{Y}_{kj} = 1 \ \forall j = 1,\ldots,n. \end{aligned} \tag{26}$$

Compared with DANMF-MRL, (26) improves multiview clustering by jointly optimizing the deep matrix representation and the latent cluster assignment, and no additional hyperparameters will be added. As a result, our proposal can make use of the latent clustering structure to generate better subspace representations, which can lead to more exact clustering assignments.

D. OS-DANMF-MRL Optimization

This section will introduce the optimization of the OS-DANMF-MRL method. In particular, the optimization algorithms about the related variables are the same as the DANMF-MRL algorithm with the exception of the representation matrix $\mathbf{H}_m^v$, the centroid matrix $\mathbf{F}$, and the cluster label matrix $\mathbf{Y}$. Therefore, the details about their update rules will be derived as follows.

1) Update Rule for Representation Matrix $\mathbf{H}_m^v$*:* Fixing other variables $\mathbf{W}_i^v$, α^v, $\mathbf{F}^v$, and $\mathbf{Y}$, the optimization problem for updating variable $\mathbf{H}_m^v$ can be rewritten as

$$\begin{aligned}\mathcal{O}(\mathbf{H}_m^v) = \operatorname{tr}\Big(&\mathbf{H}_m^{v\,T}\mathbf{H}_m^v + \mathbf{H}_m^{v\,T}\boldsymbol{\Psi}_m^{v\,T}\boldsymbol{\Psi}_m^v\mathbf{H}_m^v \\ &- 4\mathbf{X}^{v\,T}\boldsymbol{\Psi}_m^v\mathbf{H}_m^v + \lambda\mathbf{H}_m^v\mathbf{L}^v\mathbf{H}_m^{v\,T} \\ &+ (\alpha^v)^\gamma\left(\mathbf{H}_m^v - \mathbf{F}^v\mathbf{Y}\right)^T\left(\mathbf{H}_m^v - \mathbf{F}^v\mathbf{Y}\right)\Big) \\ &\text{s.t. } \mathbf{H}_m^v \geq 0. \end{aligned} \tag{27}$$

Following the similar derivation process of the update rule for $\mathbf{W}_i^v$ and $\mathbf{H}_m^v$ in Section III-B, the update rule for $\mathbf{H}_m^v$ is formulated as follows:

$$\mathbf{H}_m^v = \mathbf{H}_m^v \odot \frac{2\boldsymbol{\Psi}_m^{v\,T}\mathbf{X}^v + \lambda\mathbf{H}_m^v\mathbf{S}^v + (\alpha^v)^\gamma\mathbf{F}^v\mathbf{Y}}{\boldsymbol{\Psi}_m^{v\,T}\boldsymbol{\Psi}_m^v\mathbf{H}_m^v + \lambda\mathbf{H}_m^v\mathbf{D}^v + (1+(\alpha^v)^\gamma)\mathbf{H}_m^v}. \tag{28}$$

2) Update Rule for Centroid Matrix $\mathbf{F}^v$*:* Fixing $\mathbf{H}_m^v$ and $\mathbf{Y}$, the optimization in (26) with respect to $\mathbf{F}$ is transformed to

$$\mathcal{O}(\mathbf{F}^v) = \operatorname{Tr}\Big(\mathbf{Y}^T\mathbf{F}^{v\,T}\mathbf{F}^v\mathbf{Y} - 2\mathbf{H}_m\mathbf{Y}^T\mathbf{F}^{v\,T}\Big). \tag{29}$$

With setting the derivation with respect to $\mathbf{F}^v$ to zero, the minimum can be found when

$$\mathbf{F}^v = \mathbf{H}_m^v\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}. \tag{30}$$

3) Update Rule for Cluster Label Matrix $\mathbf{Y}$*:* Fixing $\mathbf{H}_m^v$ and $\mathbf{F}^v$, the optimization in (26) with respect to $\mathbf{Y}$ is reduced to

$$\begin{aligned}&\min_{\mathbf{Y}} \sum_{v=1}^{V}(\alpha^v)^\gamma\|\mathbf{H}_m^v - \mathbf{F}^v\mathbf{Y}\|_F^2 \\ &\text{s.t. } \mathbf{Y}_{kj} \in \{0,1\}, \quad \sum_{k=1}^{K}\mathbf{Y}_{kj} = 1 \ \forall j = 1,\ldots,n. \end{aligned} \tag{31}$$

Then, we do an exhaustive search to find out the optimal solution of (31) as

$$\mathbf{y}_j = \left\{\mathbf{e}_k \mid k = \arg\min \sum_{v=1}^{V}(\alpha^v)^\gamma\left\|\left(\mathbf{H}_m^v\right)_j - \left(\mathbf{F}^v\right)_k\right\|_F^2\right\}. \tag{32}$$

In addition, we summarize the whole alternating strategy of the OS-DANMF-MRL in Algorithm 1. Let the reconstruction error be $\mathcal{E}_i$, and we utilize the convergence rule $\mathcal{E}_i - \mathcal{E}_{i+1} \leq 10^{-4}\max(1,\mathcal{E}_i)$ to terminate the iteration when the reconstruction error between the current and prior update is sufficiently reduced.

Algorithm 1 OS-DANMF-MRL Algorithm

Input: Multi-view Data $\{\mathbf{X}^v\}_{v=1}^V$, the number of layer m, layer sizes $\{p_i\}_{i=1}^m$, hyperparameters λ.
Output: The clustering results $\mathbf{Y}$.

```
Initialization phase:
for v = 1 to V do
    Compute graph Laplacian matrix L^v from X^v;
    for i = 1 to m do
        (W_i^v, H_i^v) ← NMF(H_{i-1}^v, p_i);
    end for
end for
Fine-tuning phase:
while not converged do
    for v = 1 to V do
        for i = 1 to m do
            Update W_i^v via Eq. (15);
        end for
        Update H_m^v via Eq. (28);
        Normalize W^v and H^v by Eq. (7);
        Update α^v via Eq. (23);
        Update F^v via Eq. (30);
    end for
    for j = 1 to n do
        Update y_j via Eq. (32);
    end for
end while
```

E. Algorithm Complexity Analysis

The proposed algorithm's implementation consists of two phases, an initialization phase and a fine-tuning phase, as can be seen from Algorithm 1. For the initialization phase, the computation of graph Laplacian on all views requires $O(n^2 l^v V)$, where n is the number of samples, l^v is the number of vth view's features, and V is the number of views. In addition, both DANMF-MRL and OS-DANMF-MRL are initialized through traditional NMF layer by layer, with the complexity of order $O(t_i(nl^v p)mV)$, where t_i is the number of iterations to stop in the initialization phase, m is the number of layers, and p is the maximal layer size. For the fine-tuning phase, let k denote the average nonzero elements on each row of the graph weight matrix $\mathbf{A}$, we summarize the arithmetic operations for DANMF-MRL and OS-DANMF-MRL in Table I.

As p, k, and K are very small compared to n, let t_f be the number of iterations to stop in fine-tuning phase, the overall computational complexities for both DANMF-MRL and OS-DANMF-MRL are the same, i.e., $O(n^2kV + t_i(nl^v p)mV + t_f(l^v np)mV)$.

F. Convergence of the Algorithm

In this section, we will theoretically prove the convergence of the proposed algorithm. Note that we update one variable by fixing the others during the process of optimization. As a result, a local minimization of the objective function can be achieved if the convergence of all subproblems can be proven.

1) Convergence of the DANMF-MRL: In our proposed DANMF-MRL algorithm, since we can find the optimal solution to subproblems of α^v and $\mathbf{H}^*$, we only need to prove the convergence of update rules for $\mathbf{W}_i^v$ and $\mathbf{H}_m^v$.

Theorem 1: The objective function in (8) is nonincreasing under the update rule in (15) and (17).

To prove Theorem 1, we first give the following definition and lemma according to [29], [57], and [58].

Definition 1: If the following conditions are satisfied, then $F(\mathbf{V},\mathbf{V}^t)$ is an auxiliary function for $\mathcal{O}(\mathbf{V})$:

$$F(\mathbf{V},\mathbf{V}^t) \geq \mathcal{O}(\mathbf{V}), \quad F(\mathbf{V}^t,\mathbf{V}^t) = \mathcal{O}(\mathbf{V}^t) \tag{33}$$

where $\mathbf{V}^t$ means the value of $\mathbf{V}$ in the tth iteration.

TABLE I
FLOATING-POINT ADDITION, MULTIPLICATION, AND DIVISION FOR FINE-TUNING PHASE IN DANMF-MRL AND OS-DANMF-MRL

	Addition	Multiplication	Division	Overall
DANMF-MRL	$(5(np^2+l^v np)$ $+3(l^v p^2+p^3))mV$	$((5(np^2+l^v np)+3(l^v p^2+p^3))m$ $+(l^v+n)p+n(k+1)p)V$	$(l^v+n)pV$	$O((l^v np)mV)$
OS-DANMF-MRL	$((5(np^2+l^v np)$ $+3(l^v p^2+p^3))m+4nK^2)V$	$((5(np^2+l^v np)+3(l^v p^2+p^3))m$ $+4nK^2+(l^v+n)p+n(k+1)p)V$	$(l^v+n)pV$	$O((l^v np)mV)$

Lemma 1: If F is an auxiliary function, then $\mathcal{O}(\mathbf{V})$ is nonincreasing under the update rule

$$\mathbf{V}^{t+1} = \arg\min_{\mathbf{V}} F(\mathbf{V}, \mathbf{V}^t). \tag{34}$$

Based on Lemma 1, finding a suitable auxiliary function for the proposed lost function is essential to proving Theorem 1. Then, we provide an existing notion that will be used to define the auxiliary function according to [24].

Proposition 1: For any matrices $\mathbf{A} \in \mathbb{R}^{m\times m}$, $\mathbf{A} \in \mathbb{R}^{r\times r}$, $\mathbf{V} \in \mathbb{R}^{m\times r}$, and $\widehat{\mathbf{V}} \in \mathbb{R}^{m\times r}$, $\mathbf{A} \geq 0$, $\mathbf{B} \geq 0$, $\mathbf{V} \geq 0$, and $\widehat{\mathbf{V}} \geq 0$, it holds that

$$\sum_{j,q} \frac{(\mathbf{AVB})_{jq}\widehat{v}_{jq}^2}{v_{jq}} \geq \mathrm{Tr}\big(\widehat{\mathbf{V}}^T\mathbf{A}\widehat{\mathbf{V}}\mathbf{B}\big). \tag{35}$$

Proof of Theorem 1: To prove the convergence of $\mathbf{W}_i^v$ and $\mathbf{H}_m^v$, we define $\mathcal{O}\big(\mathbf{W}_i^v\big)$ and $\mathcal{O}\big(\mathbf{H}_m^v\big)$ be the parts of $\mathcal{O}$ containing $\mathbf{W}_i^v$ [i.e., (10)] and $\mathbf{H}_m^v$ [i.e., (16)], respectively. The first derivatives of $\mathcal{O}\big(\mathbf{W}_i^v\big)$ and $\mathcal{O}\big(\mathbf{H}_m^v\big)$ are

$$\begin{aligned}\mathcal{O}'\big(\mathbf{W}_i^v\big) &= -4{\boldsymbol{\Psi}_{i-1}^v}^T\mathbf{X}^v{\mathbf{H}_m^v}^T{\boldsymbol{\Phi}_{i+1}^v}^T \\ &\quad + 2{\boldsymbol{\Psi}_{i-1}^v}^T\boldsymbol{\Psi}_{i-1}^v\mathbf{W}_i^v\boldsymbol{\Phi}_{i+1}^v\mathbf{H}_m^v{\mathbf{H}_m^v}^T{\boldsymbol{\Phi}_{i+1}^v}^T \\ &\quad + 2{\boldsymbol{\Psi}_{i-1}^v}^T\mathbf{X}^v\mathbf{X}^{vT}\boldsymbol{\Psi}_{i-1}^v\mathbf{W}_i^v\boldsymbol{\Phi}_{i+1}^v{\boldsymbol{\Phi}_{i+1}^v}^T \qquad (36)\\ \mathcal{O}'\big(\mathbf{H}_m^v\big) &= 2\mathbf{H}_m^v + 2{\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v\mathbf{H}_m^v - 4{\boldsymbol{\Psi}_m^v}^T\mathbf{X}^v \\ &\quad + \lambda 2\mathbf{H}_m^v(\mathbf{D}^v - \mathbf{S}^v) + 2(\alpha^v)^\gamma\big(\mathbf{H}_m^v - \mathbf{H}^*\big). \qquad (37)\end{aligned}$$

Also, the second derivatives of $\mathcal{O}\big(\mathbf{W}_i^v\big)$ and $\mathcal{O}\big(\mathbf{H}_m^v\big)$ are

$$\begin{aligned}\mathcal{O}''\big(\mathbf{W}_i^v\big) &= 2{\boldsymbol{\Psi}_{i-1}^v}^T\boldsymbol{\Psi}_{i-1}^v\boldsymbol{\Phi}_{i+1}^v\mathbf{H}_m^v{\mathbf{H}_m^v}^T{\boldsymbol{\Phi}_{i+1}^v}^T \\ &\quad + 2{\boldsymbol{\Psi}_{i-1}^v}^T\mathbf{X}^v\mathbf{X}^{vT}\boldsymbol{\Psi}_{i-1}^v\boldsymbol{\Phi}_{i+1}^v{\boldsymbol{\Phi}_{i+1}^v}^T \qquad (38)\\ \mathcal{O}''\big(\mathbf{H}_m^v\big) &= 2\mathbf{I} + 2{\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v + 2\lambda(\mathbf{D}^v - \mathbf{S}^v) + 2(\alpha^v)^\gamma\mathbf{I}. \qquad (39)\end{aligned}$$

Following Definition 1 and Lemma 1, an auxiliary function $F(\mathbf{H}_m^v, {\mathbf{H}_m^v}^t)$ is constructed for $\mathcal{O}(\mathbf{H}_m^v)$ as follows:

$$\begin{aligned}F\big(\mathbf{H}_m^v, {\mathbf{H}_m^v}^t\big) &= \mathcal{O}\big(\mathbf{H}_m^v\big) + \sum_{j,q}\mathcal{O}'\big(\mathbf{H}_m^v\big)_{jq}\big(\mathbf{H}_m^v - {\mathbf{H}_m^v}^t\big)_{jq} \\ &\quad + \sum_{j,q}\frac{\big(\xi^v + \lambda\mathbf{H}_m^v\mathbf{D}^v\big)_{jq}}{{\mathbf{H}_m^v}^t{}_{jq}}\big(\mathbf{H}_m^v - {\mathbf{H}_m^v}^t\big)_{jq}^2\end{aligned} \tag{40}$$

where $\xi^v = ((1+(\alpha^v)^\gamma)\mathbf{I} + {\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v)\mathbf{H}_m^v$.

Since $F(\mathbf{H}_m^v, \mathbf{H}_m^v) = \mathcal{O}(\mathbf{H}_m^v)$ is obvious according to Definition 1, we only need to show $F(\mathbf{H}_m^v, {\mathbf{H}_m^v}^t) \geq \mathcal{O}(\mathbf{H}_m^v)$. Thus, a second-order Taylor series expansion $\mathcal{O}(\mathbf{H}_m^v)$ is performed

$$\begin{aligned}\mathcal{O}\big(\mathbf{H}_m^v\big) &= \mathcal{O}\big({\mathbf{H}_m^v}^t\big) + \sum_{j,q}\mathcal{O}'\big({\mathbf{H}_m^v}^t\big)_{jq}\big(\mathbf{H}_m^v - {\mathbf{H}_m^v}^t\big)_{jq} \\ &\quad + \sum_{j,q}\frac{\mathcal{O}''\big({\mathbf{H}_m^v}^t\big)_{jq}}{2}\big(\mathbf{H}_m^v - {\mathbf{H}_m^v}^t\big)_{jq}^2.\end{aligned} \tag{41}$$

It can be found that the formula $F(\mathbf{H}_m^v, {\mathbf{H}_m^v}^t) \geq \mathcal{O}(\mathbf{H}_m^v)$ holds only when (42) is true

$$\frac{\big(\xi^v + \lambda\mathbf{H}_m^v\mathbf{D}^v\big)_{jq}}{{\mathbf{H}_m^v}^t{}_{jq}} \geq \frac{\mathcal{O}''\big({\mathbf{H}_m^v}^t\big)_{jq}}{2}. \tag{42}$$

Since we have

$$\begin{aligned}&\Big(\Big((1+(\alpha^v)^\gamma)\mathbf{I} + {\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v\Big)\mathbf{H}_m^v\Big)_{jq} \\ &\quad = \sum_{l=1}^{p}\Big((1+(\alpha^v)^\gamma)\mathbf{I} + {\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v\Big)_{jl}\big(\mathbf{H}_m^v\big)_{lq} \\ &\quad \geq \Big((1+(\alpha^v)^\gamma)\mathbf{I} + {\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v\Big)_{jj}\big(\mathbf{H}_m^v\big)_{jq} \qquad (43)\end{aligned}$$

$$\big(\lambda\mathbf{H}_m^v\mathbf{D}^v\big)_{jq} = \sum_{s=1}^{n}\big(\mathbf{H}_m^v\big)_{js}\big(\mathbf{D}^v\big)_{sq} \geq \big(\mathbf{H}_m^v\big)_{jq}\big(\mathbf{D}^v - \mathbf{S}^v\big)_{qq}. \tag{44}$$

Thus, we can conclude that $F(\mathbf{H}_m^v, {\mathbf{H}_m^v}^t) \geq \mathcal{O}(\mathbf{H}_m^v)$ and $F(\mathbf{H}_m^v, {\mathbf{H}_m^v}^t)$ is an auxiliary function for $\mathcal{O}(\mathbf{H}_m^v)$. We can obtain the update rule for $\mathbf{H}_m^v$ by minimizing the auxiliary function for $\mathcal{O}(\mathbf{H}_m^v)$ according to Lemma 1. Since $F(\mathbf{H}_m^v, {\mathbf{H}_m^v}^t)$ is a convex quadratic function, its optimal solution is obtained

$$\begin{aligned}\mathbf{H}_m^v &= \big(\mathbf{H}_m^v\big)_{jq} - \big(\mathbf{H}_m^v\big)_{jq}\frac{\mathcal{O}'\Big(\big(\mathbf{H}_m^v\big)_{jq}\Big)}{2\big(\xi^v + \lambda\mathbf{H}_m^v\mathbf{D}^v\big)_{jq}} \\ &= \big(\mathbf{H}_m^v\big)_{jq}\frac{\big(2{\boldsymbol{\Psi}_m^v}^T\mathbf{X}^v + \lambda\mathbf{H}_m^v\mathbf{S}^v + (\alpha^v)^\gamma\mathbf{H}^*\big)_{jq}}{\big(\xi^v + \lambda\mathbf{H}_m^v\mathbf{D}^v\big)_{jq}}.\end{aligned} \tag{45}$$

The value of the cost function is nonincreasing under the above rule. Similarly, the update rules for $\mathbf{W}_i^v$ can be obtained

$$\begin{aligned}\mathbf{W}_i^v &= (\mathbf{W}_i^v)_{jq} - \big(\mathbf{W}_i^v\big)_{jq}\frac{\mathcal{O}'\Big(\big(\mathbf{W}_i^v\big)_{jq}\Big)}{2\big(\boldsymbol{\Pi}_i^v\big)_{jq}} \\ &= \big(\mathbf{W}_i^v\big)_{jq}\frac{\big(2{\boldsymbol{\Psi}_{i-1}^v}^T\mathbf{X}^v{\mathbf{H}_m^v}^T{\boldsymbol{\Phi}_{i+1}^v}^T\big)_{jq}}{\big(\boldsymbol{\Pi}_i^v\big)_{jq}}\end{aligned} \tag{46}$$

where

$$\begin{aligned}\boldsymbol{\Pi}_i^v &= {\boldsymbol{\Psi}_{i-1}^v}^T\boldsymbol{\Psi}_{i-1}^v\mathbf{W}_i^v\boldsymbol{\Phi}_{i+1}^v\mathbf{H}_m^v{\mathbf{H}_m^v}^T{\boldsymbol{\Phi}_{i+1}^v}^T \\ &\quad + {\boldsymbol{\Psi}_{i-1}^v}^T\mathbf{X}^v\mathbf{X}^{vT}\boldsymbol{\Psi}_{i-1}^v\mathbf{W}_i^v\boldsymbol{\Phi}_{i+1}^v{\boldsymbol{\Phi}_{i+1}^v}^T.\end{aligned} \tag{47}$$

It can be observed that the two update rules in (17) and (15) are equivalent to (45) and (46), respectively. Therefore, Theorem 1 is proved. ■

2) Convergence of the OS-DANMF-MRL: We just need to prove the convergence of update rules for $\mathbf{H}_m^v$ in our presented OS-DANMF-MRL algorithm since we can obtain the optimal solution to subproblems α^v, $\mathbf{F}^v$ and $\mathbf{Y}$, and the proof of $\mathbf{W}_i^v$ is the same as in the DANMF-MRL.

Theorem 2: The objective function in (26) is nonincreasing under the update rule in (28).

Proof of Theorem 2: To prove the convergence of $\mathbf{H}_m^v$, let $\mathcal{O}(\mathbf{H}_m^v)$ denote the part of $\mathcal{O}$ containing $\mathbf{H}_m^v$ [i.e., (27)]. The first and second derivatives of $\mathcal{O}(\mathbf{H}_m^v)$ are

$$\mathcal{O}'(\mathbf{H}_m^v) = 2\mathbf{H}_m^v + 2{\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v\mathbf{H}_m^v - 4{\boldsymbol{\Psi}_m^v}^T\mathbf{X}^v + \lambda 2\mathbf{H}_m^v(\mathbf{D}^v - \mathbf{S}^v) + 2(\alpha^v)^\gamma(\mathbf{H}_m^v - \mathbf{F}^v\mathbf{Y}) \quad (48)$$

$$\mathcal{O}''(\mathbf{H}_m^v) = 2\mathbf{I} + 2{\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v + 2\lambda(\mathbf{D}^v - \mathbf{S}^v) + 2(\alpha^v)^\gamma\mathbf{I}. \quad (49)$$

Similar to DANMF-MRL, we prove that each $\mathbf{H}_m^v$ is nonincreasing under the update rule (28) based on an auxiliary function as follows:

$$F(\mathbf{H}_m^v, \mathbf{H}_m^{v\,t}) = \mathcal{O}(\mathbf{H}_m^v) + \sum_{j,q}\mathcal{O}'(\mathbf{H}_m^v)_{jq}(\mathbf{H}_m^v - \mathbf{H}_m^{v\,t})_{jq} + \sum_{j,q}\frac{(\delta^v + \lambda\mathbf{H}_m^v\mathbf{D}^v)_{jq}}{\mathbf{H}_{m\,jq}^{v\,t}}(\mathbf{H}_m^v - \mathbf{H}_m^{v\,t})_{jq}^2 \quad (50)$$

where $\delta^v = ((1 + (\alpha^v)^\gamma)\mathbf{I} + {\boldsymbol{\Psi}_m^v}^T\boldsymbol{\Psi}_m^v)\mathbf{H}_m^v$. Then, the Taylor series expansion of $\mathcal{O}(\mathbf{H}_m^v)$ can be derived as

$$\mathcal{O}(\mathbf{H}_m^v) = \mathcal{O}(\mathbf{H}_m^{v\,t}) + \sum_{j,q}\mathcal{O}'(\mathbf{H}_m^{v\,t})_{jq}(\mathbf{H}_m^v - \mathbf{H}_m^{v\,t})_{jq} + \sum_{j,q}\frac{\mathcal{O}''(\mathbf{H}_m^{v\,t})_{jq}}{2}(\mathbf{H}_m^v - \mathbf{H}_m^{v\,t})_{jq}^2. \quad (51)$$

Similar to the proof in (42)–(44), we can conclude that $F(\mathbf{H}_m^v, \mathbf{H}_m^{v\,t})$ is an auxiliary function for $\mathcal{O}(\mathbf{H}_m^v)$. Accordingly, we can obtain the update rule for the $\mathbf{H}_m^v$ as

$$\mathbf{H}_m^v = (\mathbf{H}_m^v)_{jq} - (\mathbf{H}_m^v)_{jq}\frac{\mathcal{O}'\left((\mathbf{H}_m^v)_{jq}\right)}{2(\delta^v + \lambda\mathbf{H}_m^v\mathbf{D}^v)_{jq}} = (\mathbf{H}_m^v)_{jq}\frac{(2{\boldsymbol{\Psi}_m^v}^T\mathbf{X}^v + \lambda\mathbf{H}_m^v\mathbf{S}^v + (\alpha^v)^\gamma\mathbf{F}^v\mathbf{Y})_{jq}}{(\delta^v + \lambda\mathbf{H}_m^v\mathbf{D}^v)_{jq}}. \quad (52)$$

Obviously, the update rule regarding $\mathbf{H}_m^v$ is the same as rule in (28), and thus, Theorem 2 is proved. ■

IV. Experiments

In this section, we first describe the used datasets, evaluation metrics, and compared methods. Second, we analyze the parameter sensitivity of our methods. Third, we give the detailed clustering experimental results and several key observations, where the datasets with different scales are tested. Finally, we discuss the convergence of two methods.

A. Description of Datasets

In this section, we select five benchmark multiview datasets that are widely used for our experiments, including the following.

TABLE II
Statistics of Datasets Used in Experiments

Types	Names	Sizes	Classes	Views
documents	Washington	230	5	2
	BBCSports	544	5	2
images	MSRCV1	210	7	5
	Leaves	1600	100	3
	ALOI	11025	100	4

1) *Washington*[2]*:* There are 230 web documents of five categories, i.e., student, project, course, staff, and faculty. Each document is described by two views, i.e., the content of the document (1703 dimension) and the link of the same document (690 dimension).
2) *BBCSports*[3]*:* This dataset contains 544 documents samples from the BBC Sport web in five topical areas: rugby, tennis, cricket, football, and athletics. Each document is split into two related segments as our views, and the feature dimensions of different views are 3183 and 3203.
3) *MSRCV1*[4]*:* This is a scene image set, which contains seven categories, i.e., tree, building, airplane, cow, face, car, and bicycle. Each image is represented by five kinds of features: color moment (CM) with 24 dimensions, histogram of oriented gradient with 576 dimensions, GIST with 512 dimensions, local binary pattern with 256 dimensions, and centrist feature with 254 dimensions.
4) *Leaves*[5]*:* It contains 1600 samples from 100 plant species. The shape descriptor, fine scale margin, and texture histogram are used as three types of features, all with 64 dimensions.
5) *ALOI*[6]*:* This is an image dataset consisting of 11 025 samples of 100 objects. Each sample is represented by four types of features, i.e., RGB, HSV, color similarity, and Haralick. All datasets' statistics are listed in Table II.

B. Competitors and Experimental Setting

For a comprehensive comparison, we compare the proposed DANMF-MRL with the following methods.

1) *Shallow Matrix Factorization-Based Methods:* We compareour methods with five shallow matrix factorization-based models, including NMF [29], NSED [23], MultiNMF [15], NMFCC [16],[7] Consensus and Complementary information for MULTIVIEW data (2CMV) [59].[8]
2) *Deep Learning-Based Methods:* Five deep matrix factorization-based methods, including DMF [48],[9]

[2]Dataset available: https://lig-membres.imag.fr/grimal/data.html
[3]Dataset available: http://mlg.ucd.ie/datasets/segment.html
[4]Dataset available: http://research.microsoft.com/en-us/projects/objectclassrecognition/
[5]Dataset available: https://archive.ics.uci.edu/ml/datasets/One-hundred+plant+species+leaves+data+set
[6]Dataset available: http://elki.dbs.ifi.lmu.de/wiki/DataSets/MultiView
[7]Code available: https://github.com/liangnaiyao/NMFCC
[8]Code available: https://github.com/khanhluongds/Multi-view-Clustering-2CMV
[9]Code available: https://github.com/libertyhhn/DeepMF

DANMF [26],[10] DMVC [50],[11] PSDMF [52],[12] and multiview clustering via deep concept factorization (MCDCF) [54],[13] are adopted. Two recent deep autoencoder neural network-based methods, including autoencoder in autoencoder networks (AE^2-Nets) [20][14] and robust multiview clustering with incomplete information (SURE) [21].[15] Two convolutional neural network (CNN)-based methods, including a CNN multiview clustering (CNN-MVC) baseline and a convolutional autoencoder network-based method, called deep embedded multiview clustering (DEMVC) [60].[16]

Since NMF, NSED, DMF, and DANMF are single-view methods, we first concatenate the features from multiple views and then use them for concatenated data representation learning, as in [15]. Furthermore, NMFCC, 2CMV, DMVC, PSDMF, MCDCF, and ours all need the graph Laplacian matrix to be constructed using the k-NN method, where k is set to 5, as suggested in [54]. In addition, as AE^2-Nets and SURE can only handle two-view data, we report their results based on the two best data views. As for two CNN-based approaches, we follow parameter settings in [60], the 1-D convolutional kernel sizes are set to 5-5-3 with a stride of 2 by default, and the number of channels increases from 32 to 64 and 128. Eventually, the dimensionality is reduced to 10 through the use of an embedding layer comprised of ten neurons that are fully connected.

To eliminate randomness, we run each comparative algorithm ten times and report the averages and standard deviation. To facilitate readers' intuitive understanding, all numerical results are displayed in the form of percentage. The experimental environment is implemented on a server with Intel[17] Xeon[17] E5-2640 @2.40-GHz CPU and 128-GB RAM, under the Linux operating system. Also, the source code can be downloaded at: https://github.com/libertyhhn/DANMF-MRL.

Evaluation Metrics: The basic evaluation measures, accuracy (ACC), normalized mutual information (NMI), Purity, adjusted Rand (AR) index, F-score, precision, and recall, are used in this article to generally evaluate the performance of clustering. Details of these evaluation metrics can be found in [61], and higher values of metrics imply better performance for all algorithms.

C. Parameter Sensitivity Analysis

1) Influence of the Number of Layers: To investigate the influence of a different number of layers adopted by the proposed methods, we conduct numerical experiments on all five datasets, varying the models' number of layers, as reported in Figs. 3 and 4. The experiments are carried out by changing the number of layers m from 1 to 4, and the corresponding layer size is set as [50], [50, 100], [50, 100, 200], and [50, 100, 200, 300], respectively. From Figs. 3 and 4, we can see that the multilayer methods perform better than the single-layer method on all datasets, whether in ACC or NMI. This indicates that deep models may discover the hidden hierarchical information of data, so it has stronger representation ability than the shallow models. At the same time, we can find that when the depth of the models increases, the performance will initially increase until the peak performance is reached. After that, due to overfitting, when the depth of the model further increases, the performance will decrease. Thus, we need to select the appropriate number of layers for different datasets. More specifically, we adopt a two-layer DANMF-MRL structure for BBCSports, MSRCV1, and Leaves, a three-layer structure for Washington, and a four-layer structure for ALOI datasets. Also, for the OS-DANMF-MRL, we configure a two-layer structure for the Washington, BBCSports, MSRCV1, and ALOI, and a three-layer structure for Leaves datasets.

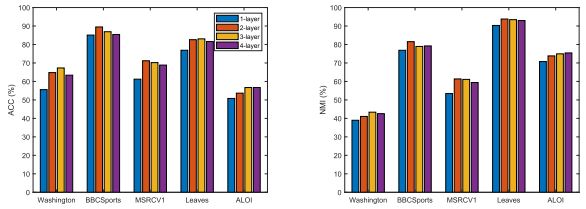


Fig. 3. Clustering results ACC and NMI of DANMF-MRL with a different number of layers on five datasets.

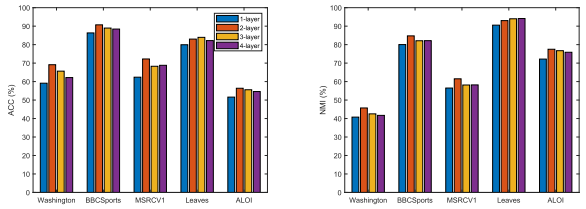


Fig. 4. Clustering results ACC and NMI of OS-DANMF-MRL with a different number of layers on five datasets.

2) Influence of the Graph Regularization and Weights Distribution: To evaluate the influence of different components in the proposed models, we investigate two important hyperparameters λ and γ. Specifically, λ and γ control the contribution of the graph regularization and distribution for the fusion weights α on different views, respectively. λ is chosen from the range $\{0.01, 0.1, 0.1, 10, 100, 1000\}$ and γ is chosen from the range $\{1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ according to the grid search strategy. Taking a document dataset Washington and an image dataset ALOI as examples, as shown in Fig. 5, we can observe that DANMF-MRL is sensitive to the change of λ but is relatively stable with varying γ, which shows the importance of considering graph regularization. From Fig. 6, we can find that when λ is set to a relatively small value and γ is set to a relatively large value, OS-DANMF-MRL can achieve better performance in most cases. This further verified that the performance of OS-DANMF-MRL is relatively robust to hyperparameters tuning.

D. Experimental Results and Analysis

In order to demonstrate the superiority of DANMF-MRL and OS-DANMF-MRL over the comparison techniques,

[10]Code available: https://github.com/smartyfh/DANMF
[11]Code available: https://github.com/hdzhao/DMF_MVC
[12]Code available: https://github.com/libertyhhn/PartiallySharedDMF
[13]Code available: https://github.com/AeroAsukara/Multi-view-clustering-via-deep-concept-factorization
[14]Code available: https://github.com/willow617/AE2-Nets
[15]Code available: https://github.com/XLearning-SCU/2022-TPAMI-SURE
[16]Code available: https://github.com/JieXuUESTC/DEMVC
[17]Registered trademark.

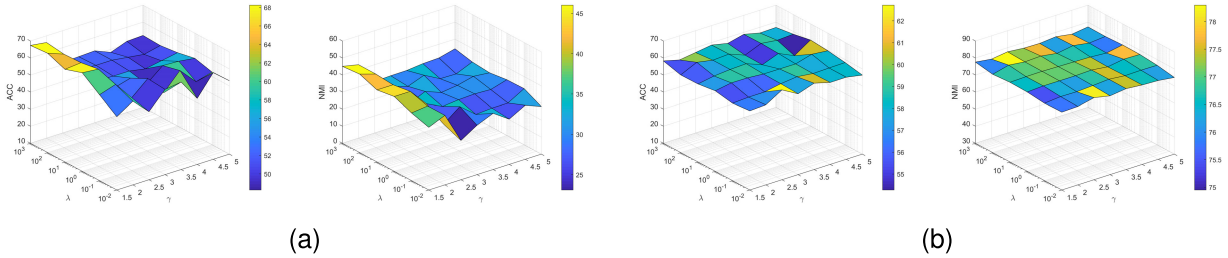


Fig. 5. ACC and NMI versus parameters λ and γ of DANMF-MRL on (a) Washington dataset and (b) ALOI dataset.

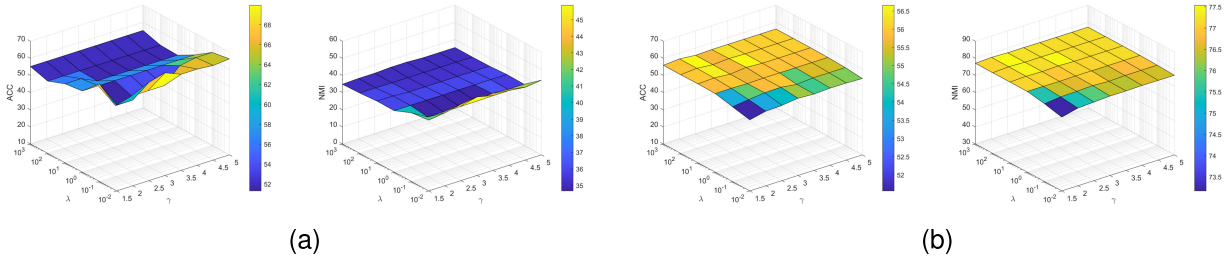


Fig. 6. (a) ACC and NMI versus parameters λ and γ of OS-DANMF-MRL on (a) Washington dataset and (b) ALOI dataset.

TABLE III

CLUSTERING PERFORMANCE ON THE WASHINGTON DATASET (MEAN ± STANDARD DEVIATION %)

Methods	ACC	NMI	Purity	AR	F-score	Precision	Recall
NMF	41.30±0.00	10.38±0.00	49.13±0.00	0.61±0.00	43.61±0.00	32.20±0.00	67.53±0.00
NSED	59.04±1.49	28.04±5.77	67.48±2.74	29.58±4.64	51.98±2.05	52.76±5.38	51.86±4.55
MultiNMF	56.01±5.16	21.32±2.74	62.57±2.35	21.76±5.64	49.68±3.53	44.84±2.83	56.56±4.00
NMFCC	62.87±2.01	34.56±3.81	69.39±1.59	39.36±3.94	57.59±2.68	61.23±3.12	54.38±2.49
2CMV	59.91±2.11	27.71±3.73	65.24±1.65	30.99±2.42	52.22±2.00	54.29±2.32	50.52±3.96
DMF	50.78±2.74	25.16±0.71	68.26±0.00	24.31±1.27	45.36±1.82	52.09±0.17	40.24±2.90
DANMF	61.39±2.94	35.50±1.51	69.22±1.46	36.92±3.16	55.70±2.76	59.77±2.32	52.30±4.19
DMVC	46.26±0.24	4.34±0.58	47.57±0.24	0.86±0.54	48.45±0.28	32.24±0.19	**97.44±0.57**
PSDMF	67.22±1.33	41.06±1.52	72.52±1.17	44.24±2.84	62.54±1.75	60.91±2.22	64.29±1.35
MCDCF	53.83±4.46	25.64±2.80	67.09±1.90	27.86±3.27	48.80±3.24	53.62±1.74	44.97±4.96
AE^2-Nets	29.18±1.09	1.97±0.29	46.56±0.14	0.10±0.34	22.48±1.05	23.60±0.94	25.19±1.85
SURE	37.30±3.87	7.54±2.52	50.65±2.65	4.26±3.98	27.22±1.90	29.47±2.70	29.32±2.25
CNN-MVC	42.70±4.03	6.12±1.66	49.13±1.87	3.12±1.49	24.19±2.54	30.14±6.20	26.55±2.39
DEMVC	49.22±2.28	26.64±1.08	64.13±1.93	24.51±1.18	38.17±2.02	42.98±1.55	38.52±2.49
DANMF-MRL	67.83±0.43	43.52±0.40	72.87±0.58	45.77±0.87	62.72±0.56	64.12±0.80	61.38±0.59
OS-DANMF-MRL	**69.13±0.00**	**45.72±0.00**	**74.35±0.00**	**48.36±0.00**	**64.83±0.00**	**64.92±0.00**	64.75±0.00

TABLE IV

CLUSTERING PERFORMANCE ON THE BBCSPORTS DATASET (MEAN ± STANDARD DEVIATION %)

Methods	ACC	NMI	Purity	AR	F-score	Precision	Recall
NMF	35.15±0.00	7.14±0.00	40.37±0.00	3.77±0.00	27.37±0.00	26.62±0.00	28.37±0.00
NSED	76.10±2.41	74.78±4.36	83.64±2.27	65.59±8.45	74.77±5.84	67.99±9.27	83.84±4.47
MultiNMF	43.83±1.38	26.40±2.54	48.13±1.08	7.64±1.81	40.32±0.97	27.53±1.00	78.25±2.33
NMFCC	81.65±7.32	70.35±8.30	82.17±6.19	70.40±11.11	77.72±8.58	74.39±6.99	81.47±10.66
2CMV	37.23±1.38	5.25±0.79	37.94±1.14	1.61±0.74	38.44±1.19	24.51±0.35	90.96±12.42
DMF	48.24±1.36	26.42±1.08	55.11±0.63	15.09±0.75	37.13±1.38	34.24±1.51	41.24±6.64
DANMF	79.45±5.47	70.42±7.59	81.29±5.72	65.49±13.17	74.19±9.95	70.15±9.55	78.98±11.57
DMVC	49.82±1.86	41.70±0.81	56.25±0.63	13.96±0.41	42.53±1.65	30.55±0.30	69.97±3.93
PSDMF	88.38±8.39	80.31±6.99	89.78±6.04	79.31±10.43	84.54±7.62	81.34±10.29	88.25±4.54
MCDCF	84.19±9.98	80.20±6.75	87.68±7.00	78.70±11.78	84.22±8.23	80.81±12.08	88.76±3.85
AE^2-Nets	27.44±1.46	1.42±0.28	35.90±0.23	0.12±0.27	21.74±0.75	24.57±1.57	23.27±1.17
SURE	29.85±2.55	3.27±0.76	37.11±1.60	0.81±1.38	24.55±2.46	26.46±2.94	24.77±2.40
CNN-MVC	31.62±2.01	5.03±1.74	38.11±2.76	2.73±1.78	27.54±3.33	29.17±3.02	29.05±3.36
DEMVC	44.25±3.00	29.59±5.59	48.95±2.77	20.52±5.93	36.74±6.96	37.74±7.39	43.98±6.46
DANMF-MRL	90.26±4.77	81.60±4.91	90.26±4.77	81.52±7.59	86.10±5.56	84.16±8.57	88.44±3.65
OS-DANMF-MRL	**90.96±0.00**	**84.28±0.00**	**90.96±0.00**	**86.36±0.00**	**89.70±0.00**	**87.42±0.00**	**92.10±0.00**

we carried out various experiments in this section. On five benchmark multiview datasets, we show extensive clustering performance in Tables III–VII, and the bold numbers in each column show the performance that is greatest for the corresponding dataset. In Table VII, "N/A" indicates that the MCDCF encountered an out-of-memory issue on

TABLE V
CLUSTERING PERFORMANCE ON THE MSRCV1 DATASET (MEAN ± STANDARD DEVIATION %)

Methods	ACC	NMI	Purity	AR	F-score	Precision	Recall
NMF	36.19±0.00	25.32±0.00	39.52±0.00	12.00±0.00	24.90±0.00	23.56±0.00	26.40±0.00
NSED	53.71±18.81	47.73±16.45	56.67±17.15	34.11±21.65	43.79±18.39	41.86±18.81	46.02±17.92
MultiNMF	67.98±1.51	58.75±1.77	67.99±1.49	53.59±2.43	57.32±2.08	56.05±2.24	58.64±1.91
NMFCC	62.95±3.15	50.72±3.58	64.76±3.32	39.44±3.31	48.12±2.82	46.57±2.93	49.81±2.98
2CMV	66.26±5.04	58.39±3.19	68.43±3.35	49.59±4.74	55.46±3.97	52.70±4.85	58.54±3.10
DMF	49.52±3.67	41.05±2.44	55.33±3.06	27.18±3.03	37.88±2.62	35.77±2.45	40.29±3.15
DANMF	66.38±0.26	52.07±0.98	68.67±1.40	51.25±1.45	57.16±1.19	51.07±1.51	64.61±0.78
DMVC	54.38±5.12	47.78±4.97	57.43±5.64	34.14±7.83	44.27±6.18	40.36±7.28	49.31± 4.95
PSDMF	67.71±4.40	61.10±2.92	69.33±3.12	51.57±4.80	57.10±4.02	54.86±4.87	59.56±3.31
MCDCF	69.62±2.77	60.42±1.86	69.62±2.77	51.12±2.46	56.72±2.11	54.32±2.01	59.30±2.28
AE^2-Nets	72.14±2.45	60.94±2.89	72.43±2.11	52.36±3.83	**71.35±2.43**	**72.24±2.47**	**72.14±2.45**
SURE	57.95±5.53	46.64±4.17	59.71±3.77	33.48±4.67	57.68±5.77	59.48±6.02	57.95±5.53
CNN-MVC	23.33±4.65	11.22±7.60	23.52±5.04	4.68±4.78	10.39±2.70	11.19±7.78	20.77±6.41
DEMVC	50.38±4.47	46.98±3.76	52.67±4.07	31.80±5.22	49.24±3.47	50.41±3.07	50.38±4.47
DANMF-MRL	71.24±0.93	**63.47±1.19**	71.52±0.62	53.25±1.46	60.08±1.39	57.06±0.51	63.49±2.88
OS-DANMF-MRL	**72.38±0.00**	61.58±0.00	**74.29±0.00**	**53.71±0.00**	60.26±0.00	59.12±0.00	61.44±0.00

TABLE VI
CLUSTERING PERFORMANCE ON THE LEAVES DATASET (MEAN ± STANDARD DEVIATION %)

Methods	ACC	NMI	Purity	AR	F-score	Precision	Recall
NMF	47.23±2.66	71.19±1.23	50.28±2.92	32.01±2.25	32.70±2.22	30.29±2.27	35.54±2.14
NSED	62.84±1.61	82.56±1.31	66.09±1.80	51.44±2.26	51.93±2.24	47.79±2.10	56.92±3.26
MultiNMF	69.65±1.53	86.58±0.44	73.10±1.28	60.70±1.51	61.10±1.49	56.26±1.71	66.87±1.19
NMFCC	76.72±1.41	91.72±0.52	80.08±1.46	71.81±1.46	72.10±1.44	66.00±1.34	79.45±1.74
2CMV	71.07±2.76	89.07±0.72	75.25±2.02	64.74±2.51	65.11±2.48	58.56±3.14	73.36±1.59
DMF	70.60±1.30	86.83±0.30	73.96±1.01	61.44±1.44	61.83±1.42	57.07±2.18	67.50±0.60
DANMF	73.41±0.92	88.53±0.49	76.81±0.71	65.18±1.54	65.54±1.52	60.69±2.05	71.25±1.02
DMVC	64.94±0.00	85.26±0.00	69.19±0.00	55.66±0.00	56.12±0.00	50.73±0.00	62.78±0.00
PSDMF	73.21±2.35	90.72±1.00	77.45±2.08	68.09±2.63	68.43±2.60	61.05±2.69	77.84±2.54
MCDCF	77.23±1.75	92.30±0.42	81.09±1.53	71.69±2.31	71.99±2.28	63.83±3.39	82.64±0.76
AE^2-Nets	62.73±9.33	82.82±5.09	65.50±9.68	52.70±10.20	59.89±10.49	62.07±9.25	62.73±9.33
SURE	50.99±3.92	80.19±1.35	55.49±3.31	36.86±6.72	48.81±3.59	53.98±2.85	50.99±3.92
CNN-MVC	7.92±2.03	29.32±7.37	8.03±2.05	2.25±1.56	4.41±1.43	8.19±2.86	7.92±2.03
DEMVC	14.04±1.26	48.47±1.58	14.09±1.28	7.20±1.04	7.02±0.83	7.09±1.28	14.04±1.26
DANMF-MRL	83.31±0.38	93.22±0.30	85.46±0.60	77.23±0.52	77.46±0.51	72.11±0.90	83.68±0.93
OS-DANMF-MRL	**83.58±0.00**	**93.64±0.00**	**85.69±0.00**	**78.78±0.00**	**78.99±0.00**	**74.36±0.00**	**84.24±0.00**

TABLE VII
CLUSTERING PERFORMANCE ON THE ALOI DATASET (MEAN ± STANDARD DEVIATION %)

Methods	ACC	NMI	Purity	AR	F-score	Precision	Recall
NMF	23.05±0.61	38.94±0.91	25.10±0.80	13.01±0.58	13.96±0.57	12.58±0.51	15.68±0.71
NSED	35.81±2.37	57.66±2.56	37.83±2.57	23.62±2.26	24.50±2.23	21.06±2.12	29.28±2.29
MultiNMF	50.79±1.37	71.73±0.97	53.74±1.34	38.33±1.36	39.03±1.34	34.05±1.36	45.76±1.29
NMFCC	50.47±1.17	69.48±0.75	52.89±0.58	37.67±1.27	38.39±1.25	33.01±1.50	45.92±1.22
2CMV	3.28±0.45	65.27±0.66	45.48±1.34	29.85±1.36	30.73±1.32	24.35±1.52	41.72±0.88
DMF	49.65±1.29	69.08±0.77	52.54±1.11	36.87±1.63	37.60±1.59	32.20±2.14	45.28±1.18
DANMF	42.09±1.79	63.78±1.09	44.22±1.51	29.61±1.16	30.43±1.15	26.09±0.91	36.50±1.66
DMVC	43.48±3.52	62.43±2.80	45.82±3.42	28.72±3.28	29.49±3.24	26.85±2.85	32.72±3.81
PSDMF	50.41±0.99	71.65±0.73	53.30±0.70	29.52±3.24	30.52±3.15	21.39±3.05	53.88±0.58
MCDCF	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AE^2-Nets	44.77±3.79	67.18±4.04	48.78±4.13	30.58±4.05	0.01±0.01	0.01±0.01	0.02±0.02
SURE	2.96±0.55	71.48±2.13	41.66±4.44	26.40±4.90	0.66±0.57	0.84±0.67	0.84±0.83
CNN-MVC	5.72±2.92	18.74±7.78	5.86±2.94	0.90±0.56	3.51±2.38	6.77±3.08	5.71±2.92
DEMVC	22.34±1.48	48.54±1.09	23.76±1.65	12.58±1.35	16.92±1.46	18.60±1.71	22.33±1.48
DANMF-MRL	56.07±1.02	75.74±0.77	59.26±0.98	43.46±1.73	44.13±1.70	36.86±1.99	54.58±1.54
OS-DANMF-MRL	**56.47±0.00**	**77.18±0.00**	**59.66±0.00**	**43.93±0.00**	**44.60±0.00**	**37.04±0.00**	**56.47±0.00**

our device. Based on the results, we have the following observations.

1) In general, compared to most baselines, the proposed DANMF-MRL and OS-DANMF-MRL offer quite competitive and steady clustering performance. In particular, compared with the second-best approach PSDMF on the BBCSports dataset, our methods achieve improvements around 2%, 4%, and 7% in terms of ACC, NMI, and AR, respectively. For the Leaves dataset, our methods outperform the second-best approach

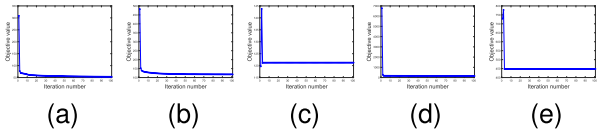


Fig. 7. Convergence curves about the proposed DANMF-MRL on five multiview datasets. (a) Washington. (b) BBCSports. (c) MSRCV1. (d) Leaves. (e) ALOI.

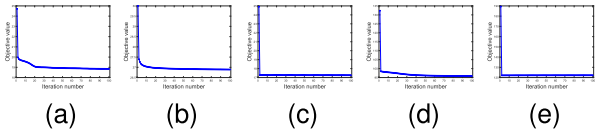


Fig. 8. Convergence curves about the proposed OS-DANMF-MRL on five multiview datasets. (a) Washington. (b) BBCSports. (c) MSRCV1. (d) Leaves. (e) ALOI.

MCDCF by 6%, 4%, and 6% in terms of average ACC, Purity, and AR. This indicates that by using deep autoencoder structures to consider the consistency and complementarity of multiple views, our approaches can learn more comprehensive representations to improve the clustering performance.

2) From a holistic perspective, OS-DANMF-MRL performs better than DANMF-MRL in most cases. Taking the Washington and MSRCV1 datasets for example, OS-DANMF-MRL achieves 1.3% and 1.14% higher than DANMF-MRL on the ACC metric. We think that the possible reason is that OS-DANMF-MRL unifies representation matrix learning and partition generation closely, which can exploit the clustering structure of multiview data effectively. Moreover, OS-DANMF-MRL has a smaller standard deviation than the method on all datasets. This is because OS-DANMF-MRL does not need to use k-means for clustering, which reduces the impact caused by the random initialization of cluster centers.
3) Our proposed methods are remarkably higher than the performance attained from three single-view approaches: 1) NMF; 2) NSED; and 3) DANMF. We think that the potential reason is that the strategy of simply concatenating multiview features into a lengthy vector ignores the fact that these views are constructed from various feature spaces with different statistical distributions.
4) On the five datasets, our methods typically outperform CNN-based techniques such as CNN-MVC and DEMVC. We believe that two factors may have contributed to this: 1) they did not consider the nonnegative structure of the data, particularly for text data (i.e., Washington and BBCsports dataset), and 2) our methods automatically fuse various views' representations via updatable weights, as opposed to their direct addition of partitions from different views.

E. Convergence Study and Running Time Analysis

According to the description in Section III-F, we have demonstrated that the updating rules for optimizing the proposed objective are convergent. Here, we experimentally examine the speed of convergence for these rules. The convergence curves about the proposed DANMF-MRL and OS-DANMF-MRL are shown in Figs. 7 and 8, respectively. The vertical axis for each figure represents the objective value, and the horizontal axis is the number of iterations. From Figs. 7 and 8, it can be observed that although DANMF-MRL is a little faster than OS-DANMF-MRL on the Washington dataset, the two methods have similar convergence speed on other datasets. In general, two methods can achieve fast convergence within about 90 iterations, further showing the rationality of our theoretical analysis.

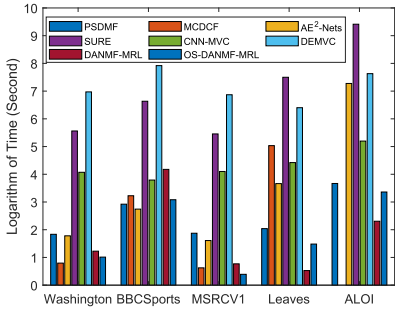


Fig. 9. Relative running time of the different approaches on five datasets. The empty bar means that the corresponding approach is out of memory on that dataset.

In order to assess the computational efficiency of the presented algorithms, Fig. 9 shows the running time of the different deep learning-based MRL algorithms on all benchmark datasets. As shown in Fig. 9, our methods outperform most approaches (AE^2-Nets, SURE, CNN-MVC, and DEMVC) in terms of running time, illustrating the proposed methods' computational effectiveness. Although MCDCF has the least running time in the Washington dataset, it is difficult to handle large-scale data problems due to its high complexity, such as ALOI dataset. As a consequence, theoretical and experimental results indicate that our methods are highly effective for MRL.

V. Conclusion

MRL is an essential task in many real-world applications. In this article, we propose a novel DANMF-MRL model to learn a comprehensive embedding representation matrix. Through the multilayer encoder and decoder, DANMF-MRL can exploit the consistency and complementary properties of multiview data in a unified optimization framework. Taking a step further, we introduce a one-step version of DANMF-MRL to address the unsatisfactory clustering performance caused by the multistage separated clustering approach. Two efficient optimization algorithms with convergence analysis are designed to solve our proposed models. Experimental comparisons on five multiview benchmark datasets demonstrate the effectiveness and superiority of our methods over state-of-the-art methods. Considering that the real-world multiview data may contain missing data in some views [62], [63], in future works, we will improve the robustness of the DANMF and extend it to tackle the incomplete multiview learning problem.

References

[1] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Inf. Fusion*, vol. 38, pp. 43–54, Nov. 2017.
[2] W. Xia, Q. Gao, Q. Wang, X. Gao, C. Ding, and D. Tao, "Tensorized bipartite graph learning for multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 5187–5202, Apr. 2023.

[3] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, Oct. 2019.

[4] K. Sun, J. Zhang, J. Liu, R. Yu, and Z. Song, "DRCNN: Dynamic routing convolutional neural network for multi-view 3D object recognition," *IEEE Trans. Image Process.*, vol. 30, pp. 868–877, 2021.

[5] X. Zhang, L. He, K. Chen, Y. Luo, J. Zhou, and F. Wang, "Multi-view graph convolutional network and its applications on neuroimage analysis for Parkinson's disease," in *Proc. AMIA Annu. Symp.*, 2018, p. 1147.

[6] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2531–2544, Dec. 2015.

[7] X. Liu, "Incomplete multiple kernel alignment maximization for clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Oct. 1, 2021, doi: 10.1109/TPAMI.2021.3116948.

[8] J. Liu, X. Liu, Y. Yang, Q. Liao, and Y. Xia, "Contrastive multi-view kernel learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 9552–9566, Aug. 2023.

[9] J. Yu, T. Zou, and G. Zhou, "Online subspace learning and imputation by tensor-ring decomposition," *Neural Netw.*, vol. 153, pp. 314–324, Sep. 2022.

[10] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Mar. 2011.

[11] Z. Kang, X. Lu, J. Yi, and Z. Xu, "Self-weighted multiple kernel learning for graph-based clustering and semi-supervised classification," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 1881–1887.

[12] J. Yu, G. Zhou, C. Li, Q. Zhao, and S. Xie, "Low tensor-ring rank completion by parallel matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3020–3033, Jul. 2021.

[13] S. Wang et al., "Fast parameter-free multi-view subspace clustering with consensus anchor guidance," *IEEE Trans. Image Process.*, vol. 31, pp. 556–568, 2022.

[14] L. Zong, X. Zhang, L. Zhao, H. Yu, and Q. Zhao, "Multi-view clustering via multi-manifold regularized non-negative matrix factorization," *Neural Netw.*, vol. 88, pp. 74–89, Apr. 2017.

[15] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. SIAM Int. Conf. Data Mining*, May 2013, pp. 252–260.

[16] N. Liang, Z. Yang, Z. Li, W. Sun, and S. Xie, "Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints," *Knowl.-Based Syst.*, vol. 194, Apr. 2020, Art. no. 105582.

[17] S. Shi, F. Nie, R. Wang, and X. Li, "Multi-view clustering via nonnegative and orthogonal graph reconstruction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 201–214, Jan. 2023.

[18] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 1083–1092.

[19] J. Xu et al., "Multi-VAE: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9214–9223.

[20] C. Zhang, Y. Geng, Z. Han, Y. Liu, H. Fu, and Q. Hu, "Autoencoder in autoencoder networks," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 15, 2022, doi: 10.1109/TNNLS.2022.3189239.

[21] M. Yang, Y. Li, P. Hu, J. Bai, J. Lv, and X. Peng, "Robust multi-view clustering with incomplete information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1055–1069, Jan. 2023.

[22] H. Huang, Y. Luo, G. Zhou, and Q. Zhao, "Multi-view data representation via deep autoencoder-like nonnegative matrix factorization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 3338–3342.

[23] B.-J. Sun, H. Shen, J. Gao, W. Ouyang, and X. Cheng, "A non-negative symmetric encoder–decoder approach for community detection," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 597–606.

[24] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2006, pp. 126–135.

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[26] F. Ye, C. Chen, and Z. Zheng, "Deep autoencoder-like nonnegative matrix factorization for community detection," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 1393–1402.

[27] X.-R. Feng, H.-C. Li, S. Liu, and H. Zhang, "Correntropy-based autoencoder-like NMF with total variation for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[28] M. Yang and S. Xu, "Orthogonal nonnegative matrix factorization using a novel deep autoencoder network," *Knowl.-Based Syst.*, vol. 227, Sep. 2021, Art. no. 107236.

[29] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 13. Cambridge, MA, USA: MIT Press, 2001, pp. 1–12.

[30] Y. Qiu, G. Zhou, Y. Wang, Y. Zhang, and S. Xie, "A generalized graph regularized non-negative tucker decomposition framework for tensor data representation," *IEEE Trans. Cybern.*, vol. 52, no. 1, pp. 594–607, Jan. 2022.

[31] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for kNN classification," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 3, pp. 1–19, 2017.

[32] Y. Yu, G. Zhou, H. Huang, S. Xie, and Q. Zhao, "A semi-supervised label-driven auto-weighted strategy for multi-view data classification," *Knowledge-Based Syst.*, vol. 255, Nov. 2022, Art. no. 109694.

[33] M. M. Kalayeh, H. Idrees, and M. Shah, "NMF-KNN: Image annotation using weighted multi-view non-negative matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 184–191.

[34] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, May 2018.

[35] Z. Jiang and X. Liu, "Adaptive KNN and graph-based auto-weighted multi-view consensus spectral learning," *Inf. Sci.*, vol. 609, pp. 1132–1146, Sep. 2022.

[36] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, pp. 1247–1255.

[37] C.-D. Wang, M.-S. Chen, L. Huang, J.-H. Lai, and P. S. Yu, "Smoothness regularized multiview subspace clustering with kernel learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 5047–5060, Nov. 2021.

[38] Y. Xie et al., "Robust kernelized multiview self-representation for subspace clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 868–881, Feb. 2021.

[39] C. Zhang et al., "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.

[40] Y. Qiu, G. Zhou, Q. Zhao, and S. Xie, "Noisy tensor completion via low-rank tensor ring," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 17, 2022, doi: 10.1109/TNNLS.2022.3181378.

[41] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, "Prior knowledge regularized multiview self-representation and its applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1325–1338, Mar. 2021.

[42] X. Zhu, S. Zhang, W. He, R. Hu, C. Lei, and P. Zhu, "One-step multi-view spectral clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 2022–2034, Oct. 2019.

[43] H. Wang, Y. Yang, and B. Liu, "GMC: Graph-based multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1116–1129, Jun. 2020.

[44] Z. Yang, N. Liang, W. Yan, Z. Li, and S. Xie, "Uniform distribution non-negative matrix factorization for multiview clustering," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3249–3262, Jun. 2021.

[45] J. Wang, F. Tian, H. Yu, C. H. Liu, K. Zhan, and X. Wang, "Diverse non-negative matrix factorization for multiview data representation," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2620–2632, Sep. 2018.

[46] H. Huang, Z. Yang, Z. Li, and W. Sun, "A converged deep graph semi-NMF algorithm for learning data representation," *Circuits Syst. Signal Process.*, vol. 2022, pp. 1–20, Jan. 2022.

[47] H. Huang, G. Zhou, N. Liang, Q. Zhao, and S. Xie, "Diverse deep matrix factorization with hypergraph regularization for multiview data representation," *IEEE/CAA J. Autom. Sinica*, early access, Oct. 4, 2022, doi: 10.1109/JAS.2022.105980.

[48] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 417–429, Mar. 2017.

[49] P. De Handschutter, N. Gillis, and X. Siebert, "A survey on deep matrix factorizations," *Comput. Sci. Rev.*, vol. 42, Jan. 2021, Art. no. 100423.

[50] H. Zhao, Z. Ding, and Y. Fu, "Multi-view clustering via deep matrix factorization," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2017, vol. 31, no. 1, pp. 1–15.

[51] S. Huang, Z. Kang, and Z. Xu, "Auto-weighted multi-view clustering via deep matrix decomposition," *Pattern Recognit.*, vol. 97, Jan. 2020, Art. no. 107015.

[52] H. Huang, N. Liang, W. Yan, Z. Yang, Z. Lit, and W. Sun, "Partially shared semi-supervised deep matrix factorization with multi-view data," in *Proc. Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2020, pp. 564–570.

[53] W. Zhao, C. Xu, Z. Guan, and Y. Liu, "Multiview concept learning via deep matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 814–825, Feb. 2021.
[54] S. Chang, J. Hu, T. Li, H. Wang, and B. Peng, "Multi-view clustering via deep concept factorization," *Knowl.-Based Syst.*, vol. 217, Apr. 2021, Art. no. 106807.
[55] K. Luong, R. Nayak, T. Balasubramaniam, and M. A. Bashar, "Multi-layer manifold learning for deep non-negative matrix factorization-based multi-view clustering," *Pattern Recognit.*, vol. 131, Nov. 2022, Art. no. 108815.
[56] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[57] Z. Li, Z. Yang, H. Zhao, and S. Xie, "Direct-optimization-based DC dictionary learning with the MCP regularizer," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3568–3579, Jun. 2021.
[58] Z. Li, Y. Li, B. Tan, S. Ding, and S. Xie, "Structured sparse coding with the group log-regularizer for key frame extraction," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 10, pp. 1818–1830, Oct. 2022.
[59] K. Luong and R. Nayak, "A novel approach to learning consensus and complementary information for multi-view data clustering," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 865–876.
[60] J. Xu, Y. Ren, G. Li, L. Pan, C. Zhu, and Z. Xu, "Deep embedded multi-view clustering with collaborative training," *Inf. Sci.*, vol. 573, pp. 279–290, Sep. 2021.
[61] L. Fu, Z. Chen, Y. Chen, and S. Wang, "Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering," *IEEE Trans. Multimedia*, early access, Jun. 24, 2022, doi: 10.1109/TMM.2022.3185886.
[62] N. Liang, Z. Yang, and S. Xie, "Incomplete multi-view clustering with sample-level auto-weighted graph fusion," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 6, pp. 6504–6511, Jun. 2023.
[63] J. Yu, H. Huang, Q. Duan, Y. Wang, and T. Zou, "Incomplete multiview clustering via low-rank tensor ring completion," *Int. J. Intell. Syst.*, vol. 2023, pp. 1–15, May 2023.

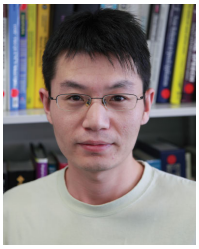
Haonan Huang (Graduate Student Member, IEEE) received the B.Eng. degree from the School of Information, Beijing Institute of Technology, Zhuhai, China, in 2018, and the M.S. degree from the Guangdong University of Technology, Guangzhou, China, in 2021, where he is currently pursuing the Ph.D. degree.

He was a Visiting Scholar with Lehigh University, Bethlehem, PA, USA, from 2022 to 2023. He has been a jointly trained Ph.D. student supported by the China Scholarship Council (CSC) at the RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan, since 2023. He has authored several peer-reviewed papers that have been published in prestigious journals and conferences, such as IEEE/CAA JOURNAL OF AUTOMATICA SINICA, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, Knowledge-Based Systems (KBS), IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and IEEE International Conference on Data Mining (ICDM).

Guoxu Zhou (Member, IEEE) received the Ph.D. degree in intelligent signal and information processing from the South China University of Technology, Guangzhou, China, in 2010.

He worked as a Research Scientist with the Brain Science Institute, RIKEN, Tokyo, Japan. He is currently a Full Professor and the Dean of the School of Automation, Guangdong University of Technology, Guangzhou. He has authored over 100 peer-reviewed articles that have been published in prestigious journals, such as the PROCEEDINGS OF THE IEEE, *IEEE Signal Processing Magazine*, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS), IEEE TRANSACTIONS ON CYBERNETICS, and IEEE TRANSACTIONS ON IMAGE PROCESSING. His current research interests include tensor analysis, intelligent information processing, and artificial intelligence.

Dr. Zhou serves as an Associate Editor for IEEE TNNLS and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS.



Qibin Zhao (Senior Member, IEEE) received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2009.

He was a Research Scientist with the Brain Science Institute, RIKEN, Tokyo, Japan, from 2009 to 2017. He is currently the Team Leader at the RIKEN Center for Advanced Intelligence Project and a Visiting Professor with the Guangdong University of Technology, Guangzhou, China. He has published more than 150 scientific articles and coauthored two monographs on tensor networks. His research interests include machine learning, tensor factorization and tensor networks, and brain signal processing.

Dr. Zhao serves as the Area Chair for ML conferences of Conference on Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), and Asian Conference on Machine Learning (ACML), and an Action Editor for *Neural Networks* and *Transactions on Machine Learning Research*.

Lifang He (Member, IEEE) received the Ph.D. degree in computer science from the South China University of Technology, Guangzhou, China, in 2014.

She is currently an Assistant Professor with the Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA. She has published more than 150 papers in refereed journals and conferences, with an H-index of 38 and more than 4000 citations. Her primary research interests focus on machine learning, artificial intelligence, and their applications in medical data mining.

Dr. He has served as a reviewer and a Program Committee Member for various esteemed journals and conferences, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Conference on Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining, hosts an influential annual conference (SIGKDD), Annual AAAI Conference on Artificial Intelligence (AAAI), and IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Furthermore, she has been serving as the Chair for the Computer Science Chapter at the IEEE Lehigh Valley Section since January 2023.

Shengli Xie (Fellow, IEEE) received the M.S. degree in mathematics from Central China Normal University, Wuhan, China, in 1992, and the Ph.D. degree in control theory and applications from the South China University of Technology, Guangzhou, China, in 1997.

He is currently the Director of the Laboratory for Intelligent Information Processing (LIIP) and a Full Professor with the Guangdong University of Technology, Guangzhou. He is also a Full Foreign Member of the Russian Academy of Engineering (RAE), Moscow, Russia. He has authored or coauthored two monographs and more than 100 scientific papers published in journals and conference proceedings. His current research interests include automatic control and signal processing, especially blind signal processing and image processing.

Dr. Xie is a fellow of the Chinese Association of Automation (CAA).