A Unified Framework Based on Graph Consensus Term for Multiview Learning

Xiangzhu Meng, Lin Feng, Chonghui Guo, Huibing Wang, and Shu Wu, *Senior Member, IEEE*

***Abstract*— In recent years, multiview learning technologies have attracted a surge of interest in the machine learning domain. However, when facing complex and diverse applications, most multiview learning methods mainly focus on specific fields rather than provide a scalable and robust proposal for different tasks. Moreover, most conventional methods used in these tasks are based on single view, which cannot be readily extended into the multiview scenario. Therefore, how to provide an efficient and scalable multiview framework is very necessary yet full of challenges. Inspired by the fact that most of the existing single view algorithms are graph-based ones to learn the complex structures within given data, this article aims at leveraging most existing graph embedding works into one formula via introducing the graph consensus term and proposes a unified and scalable multiview learning framework, termed graph consensus multiview framework (GCMF). GCMF attempts to make full advantage of graph-based works and rich information in the multiview data at the same time. On one hand, the proposed method explores the graph structure in each view independently to preserve the diversity property of graph embedding methods; on the other hand, learned graphs can be flexibly chosen to construct the graph consensus term, which can more stably explore the correlations among multiple views. To this end, GCMF can simultaneously take the diversity and complementary information among different views into consideration. To further facilitate related research, we provide an implementation of the multiview extension for locality linear embedding (LLE), named GCMF-LLE, which can be efficiently solved by applying the alternating optimization strategy. Empirical validations conducted on six benchmark datasets can show the effectiveness of our proposed method.**

***Index Terms*— Graph consensus term, iterative alternating strategy, multiview learning, unified framework.**

Manuscript received 31 January 2021; revised 8 October 2021 and 14 June 2022; accepted 16 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61972064, Grant 62002041, Grant 72001191, and Grant U19B2038; in part by the Henan Natural Science Foundation under Grant 201300410442; and in part by the Henan Philosophy and Social Science Program under Grant 2020CZH009. *(Corresponding authors: Lin Feng; Huibing Wang.)*

Xiangzhu Meng and Shu Wu are with the Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: xiangzhu.meng@cripac.ia.ac.cn; shu.wu@nlpr.ia.ac.cn).

Lin Feng is with the School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China (e-mail: fenglin@dlut.edu.cn).

Chonghui Guo is with the Institute of Systems Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: dlutguo@dlut.edu.cn).

Huibing Wang is with the College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China (e-mail: huibing.wang@dlmu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2022.3201498.

Digital Object Identifier 10.1109/TNNLS.2022.3201498

I. Introduction

With the rapid development of the information era, more and more data can be obtained from different domains or described from various perspectives, which have gained extensive attention from researchers in recent years. For examples, an image could be represented by different visual descriptors [1] to reveal its color, texture, and shape information; the document could be translated as different versions via various languages [2]. These data collected from multiple views depict different perspectives for one object, indicating that one view may contain some knowledge information that other views do not involve. A feasible manner to deal with multiview data is proposed to concatenate different views together as one view. But this way not only lacks physical meaning owing to its specific statistical property in each view, but also ignores the complementary nature of different views. Therefore, how to effectively discover the rich information of multiple views and the underlying structures within multiview data is the main challenge. To take full advantage of rich information in multiview data, various multiview learning methods [3], [4] have been well investigated in many applications (e.g., classifications [5], [6], clustering [7], [8], dimension reduction [9], [10] reidentification [11], [12], etc). Among these works, one popular class of multiview learning methods [13], [14], [15], [16] is to consider the weighted combination of different views to explore a common latent space shared by all views in integrating multiview information. For example, auto-weighted multiview graph learning (AMGL) [14] is an auto-weighted multiple graph learning method, which can automatically allocate ideal weight for each view to find common low-dimensional representations. Unlike these works above, to further guarantee the complementary effects across different views, these algorithms in co-training [17], [18] and co-regularization [19], [20] styles are developed to explore the complementary information among different views. The former iteratively maximizes the mutual agreement on different views to guarantee the consistency of different views. The latter employs co-regularization terms of discriminant functions, added into the objective function, to ensure the consensus among distinct views. However, these methods may produce unsatisfactory results when facing such multiple views that are highly related but slightly different from each other. More importantly, these above methods mainly focus on specific fields so that cannot provide a unified framework for different tasks. Even though some general

multiview frameworks [21], [22], [23] have been proposed in recent years, but these works usually tend to some specific styles of multiview models, such as multiview subspace learning. To this end, there are not still sufficient researches on generalized multiview frameworks. Inspired by graph embedding framework [24] that most of subspace learning methods [25], [26] and their kernel extensions [27], [28] could be also cast as special embedding methods based on the graph, and the fact that most existing multiview works are graph-based ones, this article attempts to handle these above issues based on graph embedding technology.

This article proposes a novel framework, named graph consensus multiview framework (GCMF), for multiview learning problems. GCMF aims to provide a scalable and robust proposal for different multiview tasks, by leveraging most existing graph embedding works based on single view into a unified formulation. Specifically, to preserve the diversity property of intrinsic information in each view, this model explores the intrinsic graph structure in each view based on single-view graph method; the graph consensus term based on learned graphs is proposed to consider the correlations among multiple views jointly, which can fully exploit the complementary information among different learned representations. For solving the proposed GCMF, this article develops a rough paradigm based on iterative alternating strategy, and the self-weighting strategy is optionally utilized in the optimization process. To facilitate related multiview researches and improve the convenience for readers, the proposed framework is utilized to implement the multiview extension of locality linear embedding (LLE) [29], named GCMF-LLE. Finally, extensive experiments based on the applications of document classification, face recognition, and image retrieval validate the ideal performance of our proposed method. The major contributions in this article can be listed as follows.

1) We propose a novel unified framework named GCMF to leverage most of existing single-view works based on the graph into a unified formula, which can be used in complex and diverse applications.
2) Graph consensus term is proposed to exploit the complementary information among different learned representations, in which the construction manner for learned graph can be flexibly chosen according to the practical tasks.
3) An implementation of the multiview extension for LLE is provided to construct a novel multiview learning method, named GCMF-LLE, which can facilitate the usage and understanding for readers.

The remainder of this article is organized as follows. In Section II, we briefly review multiview learning methods closely related to our method; in Section III, we describe the construction procedure of the proposed GCMF and its optimization algorithm; in Section IV, the proposed framework is utilized to implement the multiview extension of LLE; in Section V, extensive experiments on six datasets evaluate the effectiveness of our proposed approach; in Section VI, we make the conclusion of article.

II. Related Work

In this section, we review a brief comprehension of the related works close to the proposed method, which can be divided into three following categories.

A. CCA-Based Multiview Methods

Canonical correlation analysis (CCA) [30] and its kernel extension [31] are representative methods for cross-view features alignment. Suppose that two sets of $\boldsymbol{X}$ and $\boldsymbol{Y}$, consisting of N observations, are drawn jointly from a probability distribution. CCA-based multiview methods [21], [32], [33] employs CCA to project the two views into the common subspace by maximizing the cross correlation between two views, which can be expressed as follows:

$$\text{Corr}(\boldsymbol{X}, \boldsymbol{Y}) = \text{tr}\left(\boldsymbol{W}_X^T \boldsymbol{X} \boldsymbol{Y}^T \boldsymbol{W}_Y\right) \tag{1}$$

where $\boldsymbol{W}_X$ and $\boldsymbol{W}_Y$ denote the projecting matrix of the set $\boldsymbol{X}$ and the set $\boldsymbol{Y}$, respectively. $\text{tr}(\cdot)$ is the trace of the matrix.

In particular, CCA is further generalized in the multiview situation, named multiview CCA (MvCCA) [32], which can handle multiview data with more than two views. Multiview discriminant analysis [33] is proposed to extend LDA [28] into a multiview setting, which projects multiview features into one discriminative common subspace. Generalized multiview analysis (GMA) [21] solves a joint and relaxed problem of the form of the quadratic constrained quadratic program (QCQP) over different feature spaces to obtain a common linear subspace, which generalizes CCA for multiview scenario, i.e., cross-view classification and retrieval. Inspired by the advance of deep neural networks [34], Andrew *et al.* [35] proposed deep CCA to capture the association of high semantic level among multiview data by associating the representation among multiple views at the higher level. However, dimensionalities of different views must keep equal with each other in these CCA-based works.

B. HSIC-Based Multiview Methods

Hilbert–Schmidt independence criterion (HSIC) [36] measures dependence of the learned representations of different views by mapping variables into a reproducing kernel Hilbert space, which could be expressed as follows:

$$\text{HSIC}(\boldsymbol{X}, \boldsymbol{Y}) = (N-1)^{-2}\text{tr}(\boldsymbol{K}_X \boldsymbol{H} \boldsymbol{K}_Y \boldsymbol{H}) \tag{2}$$

where $\boldsymbol{K}_X$ and $\boldsymbol{K}_Y$ denote the Gram matrix of the set $\boldsymbol{X}$ and the set $\boldsymbol{Y}$, respectively. $\boldsymbol{H} = \boldsymbol{I} - N^{-1}\boldsymbol{1}\boldsymbol{1}^T$ centers the Gram matrix $\boldsymbol{K}_X$ or $\boldsymbol{K}_Y$ to have zero mean in the feature space.

HSIC-based multiview learning methods [37], [38], [39], [40], [41], [42], [43] explore complementary information by utilizing HSIC to measure the correlations of different views. Compared to those methods based on CCA, such HSIC-based multiview methods can relieve the restriction of equal dimensionalities for different views. Among them, the work [37] employs a kernel dependence measure of HSIC to quantify alternativeness between clustering solutions of two views, which iteratively discovers alternative clusterings. Similarly,

the work [39] exploits the complementarity information of multiple views based on HSIC to enhance the correlations (or penalize the disagreement) across different views during the dimensionality reduction, and explores the correlations within each view jointly. Latent multiview subspace clustering (LMSC) [40] is proposed to seek the underlying latent representation shared by all views, which simultaneously combines the HSIC term to discover the complementary information from multiple views. Similar to these works, similarity and diversity induced paired projection (SDPP) [42] introduces the HSIC term as a co-regularization to explicitly enforce the diversity, and removes the view-specific information that does not contribute to task. However, these HSIC-based works usually incorporate the inner product kernel to construct the HSIC term, which might lead to unsatisfactory performance when facing those nonlinear cases. Differing from those methods above, graph consensus term proposed in this article not only can overcome the limitation of dimensional equivalent across views but might be more applicable for the nonlinear cases.

C. Graph-Based Multiview Methods

Generally, most of multiview learning methods belong to the category of the graph-based method. At the aspect of graph-based methods, traditional graph-based methods mainly aim to explore the relationships among data points, and its unified form can be generally expressed as follows:

$$\min_{\boldsymbol{U}^v \in \boldsymbol{C}^v} \boldsymbol{\mathcal{F}}(\boldsymbol{G}^v, \boldsymbol{U}^v) + \lambda \boldsymbol{\Omega}(\boldsymbol{U}^v) \tag{3}$$

where $\boldsymbol{C}^v$ denotes the different constraints on the embedding $\boldsymbol{U}^v$. $\boldsymbol{\mathcal{F}}(\cdot,\cdot)$ is the loss function defined on the embedding $\boldsymbol{U}^v$ and the graph $\boldsymbol{G}^v$, and $\boldsymbol{\Omega}(\cdot)$ stands for the regularization term of the embedding $\boldsymbol{U}^v$. Graph embedding framework [24] implies that most of subspace learning methods [25], [26] and their kernel extensions [27], [28] could be also cast as special graph-based embedding methods like the form in (3). Other graph-based ones are using the so-called self-expressiveness property, and representative works include low-rank representation (LRR) [44], [45], sparse subspace learning [46], [47], etc.

On the contrary to traditional graph-based methods, graph-based multiview methods aim to exploit the intrinsic structure information within multiview data. Thereinto, the most representative group of multiview methods [8], [14], [48], [49], [50], [51], [52], [53], [54] aim to fuse multiple features or graphs into one common latent space shared by all views. Multiple kernel learning (MKL) [49], [52] is also a natural way to integrate different views based on the direct combination of different views and learn a common low-dimensional representation. Different from MKL, parameter-free multiview learning methods [14] provide a self-weighting strategy to fuse multiple graph information without additional parameters. Besides, learning a shared graph among all views is also an efficient manner to integrate the diversity information within multiview data, e.g., graph-based multiview clustering (GMC) [8] and multiview latent proximity learning (MLPL) [54]. However, these above methods do not explicitly consider the complementarity efforts across different views. Besides, these existing graph-based methods for single view are not applicable for extending to the multiview setting directly, so that we cannot take full advantage of these works. Unlike these graph-based methods, this article approximately regards most single-view methods as graph-based works and leverages most of them into a unified framework while comprehensively considering rich information within multiview data.

TABLE I
IMPORTANT NOTATIONS USED IN THIS ARTICLE

Notation	Description
M	The number of views
N	The number of samples
$\boldsymbol{X}^v$	The features set in the vth view
$\boldsymbol{x}_i^v$	The ith sample in the vth view
$\boldsymbol{K}^v$	The kernel matrix in the vth view
$\boldsymbol{U}^v$	The embedding in the vth view
$\boldsymbol{G}^v$	The graph matrix defined in original features $\boldsymbol{X}^v$
$\boldsymbol{G}_*^v$	The graph matrix defined in learnt features $\boldsymbol{U}^v$
$\boldsymbol{\mathcal{F}}(\cdot,\cdot)$	The loss function defined on the embedding $\boldsymbol{U}^v$
$\boldsymbol{\Omega}(\cdot)$	The smooth regularized term defined on the embedding $\boldsymbol{U}^v$
$Tr(\cdot)$	The trace of the matrix

III. METHODOLOGY

In this section, we discuss the intuition of our proposed framework, named GCMF. Here, we introduce the graph consensus term to regularize the dependence among different views. For clarity, the flowchart of GCMF is shown in Fig. 1. Subsequently, a rough paradigm based on iterative alternating strategy is proposed to solve the solution of GCMF. Finally, we provide a more comprehensive explanation by comparing it with other related multiview learning methods. For convenience, the important notations used in the remainder of this article are summarized in Table I.

A. Problem Definition

Given a multiview dataset consisting of M views, the data in the vth view ($1 \le v \le M$) can be denoted as $\boldsymbol{X}^v = \{\boldsymbol{x}_1^v, \boldsymbol{x}_2^v, \ldots, \boldsymbol{x}_N^v\}$, in which N is the number of samples. The proposed method aims to obtain the graph structure or the embedding in each view under the multiview setting. We separately employ $\boldsymbol{G}^v \in \mathbb{R}^{N\times N}$ and $\boldsymbol{U}^v \in \mathbb{R}^{d^v\times N}$ to denote the graph structure or the embedding in the vth view, where d^v is the dimensionality of the vth view. Differing from the graph $\boldsymbol{G}^v$ defined on $\boldsymbol{X}^v$, $\boldsymbol{G}_*^w$ is the graph constructed by the learned embedding $\boldsymbol{U}^v$. For the multiview setting, a naive way is to incorporate all views in (3) directly as follows:

$$\min_{\{\boldsymbol{U}^v \in \boldsymbol{C}^v, 1\le v\le M\}} \sum_{v=1}^{M} \boldsymbol{\mathcal{F}}(\boldsymbol{G}^v, \boldsymbol{U}^v) + \lambda \boldsymbol{\Omega}(\boldsymbol{U}^v). \tag{4}$$

Intuitively, this naive way implements graph embedding problem for each view independently and fails to exploit the diversity information of these multiple views. More importantly, this way neglects the correlations of these multiple views, so that the complementary information among multiple views cannot be made full advantage to enforce all views to learn from each other. Accordingly, how to efficiently discover the complementary information among views is the

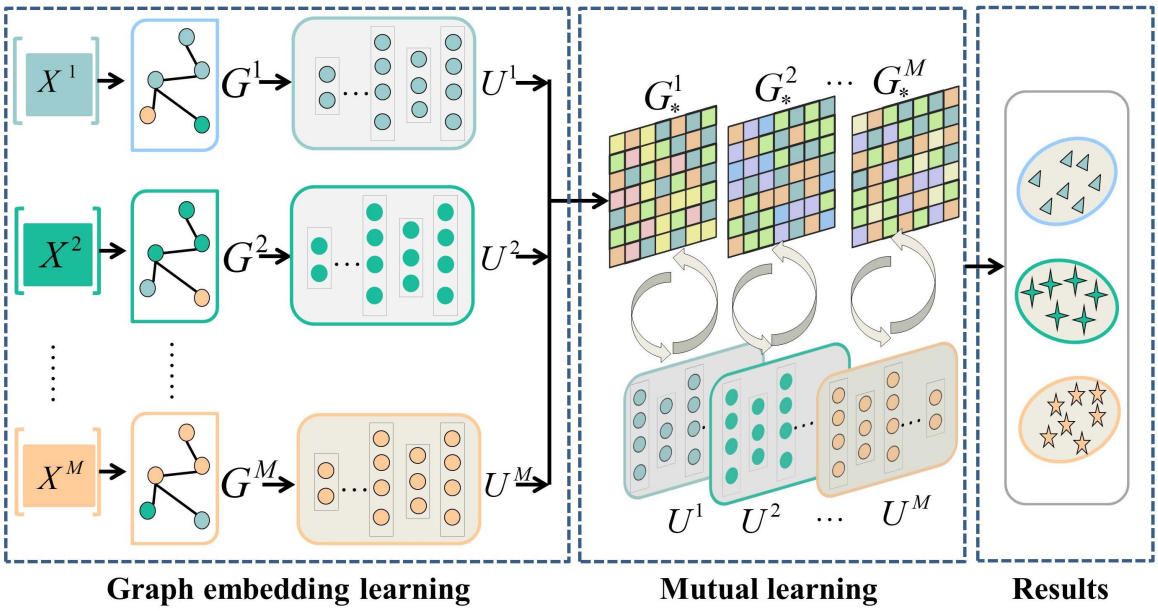


Fig. 1. Flowchart of the proposed GCMF. Given a collection of samples with M views, e.g., $\{\boldsymbol{X}^1, \boldsymbol{X}^2, \ldots, \boldsymbol{X}^M\}$. GCMF first explores the graph structure $\boldsymbol{G}^v$ in each view by graph embedding model independently. Based on the graph $\boldsymbol{G}^v$, we can initialize the embedding $\boldsymbol{U}^v$ in the vth view. Later, mutual learning based on graph consensus term $\text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w)$ is to enforce different views to learn with each other, where the graph $\boldsymbol{G}_*^w$ is built on the learned embedding $\boldsymbol{U}^v$. With the specific-view representations $\{\boldsymbol{U}^1, \boldsymbol{U}^2, \ldots, \boldsymbol{U}^M\}$ learned, the kNN classifier can be utilized to obtain the final classification results.

key point. Besides those works based on CCA or HSIC, traditional solutions usually minimize the difference between the embeddings of pairwise views directly. However, such methods are only suitable for the case that the dimensionalities are equal for different views. For these reasons, it is necessary and worthy to develop a novel co-regularization term with better scalability and robustness to enforce different views to mutually learn.

B. Graph Consensus Term

In this article, we investigate to measure the dependence among all views based on graph structures, which reveals the relationships among all samples in each view. Specifically, we attempt to construct the view-structure consensus in terms of learned graphs to regularize the dependence between two views. Taking the example with two graphs $\boldsymbol{G}_*^v$ and $\boldsymbol{G}_*^w$ in the vth view and the wth view, if $\boldsymbol{G}_*^v$ and $\boldsymbol{G}_*^w$ are obtained by the same style of graph approaches, discovering similarly property of individual view, we call such two graphs as homogeneous graphs; in contrast, if two graphs are solved by the different style of graph approaches, we call such two graphs as heterogeneous graphs. When facing the case of homogeneous graphs, directly minimizing the gap $\|\boldsymbol{G}_*^v - \boldsymbol{G}_*^w\|_F^2$ between two graphs is to make the relationships among all samples, as consistent as possible. However, the diversity information from multiple views might be reduced in this way. For the case of heterogeneous graphs, it is unsuitable to straightforward minimize $\|\boldsymbol{G}_*^v - \boldsymbol{G}_*^w\|_F^2$ owing to their different construction styles. Inspired by the property that the graph coefficients could reflect the intrinsic geometric properties of one given view, which are invariant to exactly such transformations, we expect their characterization of geometry structure in the one view to be equally valid for the other view on the manifold. That is to say, the relationship between two samples in the vth view is expected to be closer if the distance in the wth view is larger. Accordingly, we propose the following cost function as measure of dependence between two views:

$$\begin{aligned}\text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w) &= \sum_{i,j=1}^{N} \left\|\boldsymbol{U}_i^v - \boldsymbol{U}_j^v\right\|_2^2 \boldsymbol{G}_{*ij}^w \\ &= \text{tr}\left(\boldsymbol{U}^v \left(\boldsymbol{D}_*^w - \boldsymbol{G}_*^w\right) \boldsymbol{U}^{v\,T}\right) \end{aligned} \tag{5}$$

where $\boldsymbol{D}_*^w$ denote a diagonal matrix, in which the ith diagonal element in $\boldsymbol{D}_*^w$ is the sum of all elements in the ith row of $\boldsymbol{G}_*^w$.

Besides, when the graph structure specifically reflects the reconstruction relationships among samples, i.e., LRR [44], we try to solve the self-representation issue by the following form:

$$\boldsymbol{U}^v = \boldsymbol{U}^v \boldsymbol{G}_*^v + \boldsymbol{E}^v \tag{6}$$

where $\boldsymbol{E}^v$ denotes the error term of samples reconstruction. At this time, we investigate to measure the dependence between two views from the aspect of space reconstruction. That is, we expect that reconstruction relationships among samples in the one view could be equally preserved in the other view on the manifold. Therefore, we additionally could utilize the following cost function to measure the consensus between the vth view and the wth view

$$\begin{aligned}\text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w) &= \left\|\boldsymbol{U}^v - \boldsymbol{U}^v \boldsymbol{G}_*^w\right\|_F^2 \\ &= \text{tr}\left(\boldsymbol{U}^v (\boldsymbol{I}_N - \boldsymbol{G}_*^w)(\boldsymbol{I}_N - \boldsymbol{G}_*^w)^T \boldsymbol{U}^{v\,T}\right). \end{aligned} \tag{7}$$

For convenience, we can further summarize the graph consensus term into a unified form $\text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w) = \text{tr}(\boldsymbol{U}^v \boldsymbol{L}^w \boldsymbol{U}^{v\,T})$ through (5)–(7), where $\boldsymbol{L}^w$ is just dependent on the graph $\boldsymbol{G}_*^w$.

In the above discussion, we provide two formulas of $\boldsymbol{L}^w$ based on the consistent preservation between two views. To sum up, we could utilize the graph consensus term $\text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w)$ to co-regularize the dependence among different views and simultaneously obtain the graph structure or embedding for each view.

C. Multiview Learning Framework Based on Graph Consensus Term

To fully explore the correlations and complementary information among multiple views, we employ the graph consensus term in (5)–(7) to encourage the new representations of different views to be close to each other. Accordingly, by combining graph embedding loss term in each view with graph consensus term among all views, the overall objective function could be formulated as follows:

$$\min_{\{\boldsymbol{U}^v \in \boldsymbol{C}^v, \boldsymbol{G}_*^w, 1 \le v \le M\}} \underbrace{\sum_{v=1}^{M} \boldsymbol{\mathcal{F}}(\boldsymbol{G}^v, \boldsymbol{U}^v)}_{\text{Graph embedding loss}} + \underbrace{\lambda_R \sum_{v=1}^{M} \boldsymbol{\Omega}(\boldsymbol{U}^v)}_{\text{Normalization term}} + \underbrace{\lambda_C \sum_{v \ne w} \text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w)}_{\text{Graph consensus term}} \quad (8)$$

where $\lambda_R > 0$ and $\lambda_C > 0$ are two tradeoff parameters corresponding to the smooth regularized term and graph consensus term, respectively. Under the assumption that space structures in different views could reflect intrinsic properties diversely, the first term ensures that the graphs are constructed for homogeneous structures. The second term guarantees the smoothness within each view independently, and the third term enforces that the learned representations $\{\boldsymbol{U}^v, 1 \le v \le M\}$ should learn from each other to minimize the gap between them. In this way, when facing multiview issues, our proposed framework could deal with the diversity information, smooth regularized terms, and complementary information among multiple views jointly.

1) Optimization Procedure: With the alternating optimization strategy, (8) could be approximately solved. That is to say, we solve each view at a time while fixing other views. Specifically, with all views but $\boldsymbol{U}^v$ fixed, we get the following optimization problem for the vth view:

$$\min_{\boldsymbol{U}^v \in \boldsymbol{C}^v} \boldsymbol{\mathcal{F}}(\boldsymbol{G}^v, \boldsymbol{U}^v) + \lambda_R \boldsymbol{\Omega}(\boldsymbol{U}^v) + \lambda_C \sum_{1 \le v \ne w}^{M} (\text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w) + \text{Reg}(\boldsymbol{U}^w, \boldsymbol{G}_*^v)). \quad (9)$$

Note that in $\text{Reg}(\boldsymbol{U}^w, \boldsymbol{G}_*^v)$, $\boldsymbol{G}_*^v$ is dependent on the target variable $\boldsymbol{U}^v$ and (9) could not be directly solved. But if $\boldsymbol{G}_*^v$ is set to be stationary, $\text{Reg}(\boldsymbol{U}^w, \boldsymbol{G}_*^v)$ will be reduced a constant term on $\boldsymbol{U}^v$. Without considering the constant terms, (9) will reduce to the following:

$$\min_{\boldsymbol{U}^v \in \boldsymbol{C}^v} \boldsymbol{\mathcal{F}}(\boldsymbol{G}^v, \boldsymbol{U}^v) + \lambda_R \boldsymbol{\Omega}(\boldsymbol{U}^v) + \lambda_C \sum_{1 \le v \ne w}^{M} \text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w) \quad (10)$$

which looks simpler to be solved. Notably, we assign the same importance for other views in updating $\boldsymbol{U}^v$. After finely weighting other views, the performance may be further improved. For this reason, we optionally utilize the weighting strategy as follows:

$$\min_{\boldsymbol{U}^v \in \boldsymbol{C}^v} \boldsymbol{\mathcal{F}}(\boldsymbol{G}^v, \boldsymbol{U}^v) + \lambda_R \boldsymbol{\Omega}(\boldsymbol{U}^v) + \lambda_C \sum_{1 \le v \ne w}^{M} \boldsymbol{\alpha}_v^w \text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w) \quad (11)$$

where $\boldsymbol{\alpha}_v^w \le 0$ denotes the importance of the wth view in updating $\boldsymbol{U}^v$, i.e., $\sum_{v \ne w} \boldsymbol{\alpha}_v^w = 1$. Usually, we can adjust the importance parameter $\boldsymbol{\alpha}_v^w$ by grid search technology. Besides, inspired by these works [14], [20], [52], we additionally provide a self-weighting strategy to improve the efficiency of the weight assignment, which can be expressed as follows:

$$\boldsymbol{\alpha}_v^w = \frac{f(\text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w))}{\sum_{v \ne w} f(\text{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w))} \quad (12)$$

where $f(\cdot)$ is a scalar function to adjust the specific-view weight, such as exponential function.

Suppose that $\boldsymbol{U}^v$ could be effectively calculated by solving (10), this $\boldsymbol{U}^v$ could be continuously used to update $\boldsymbol{G}_*^v$ according to the construction manner of chosen learned graph method, which inspires us to compute $\boldsymbol{U}^v$ and $\boldsymbol{G}_*^v$ iteratively. The whole procedure to solve (8) is summarized in **Algorithm 1**.

Algorithm 1 Optimization for GCMF

Input: The multiview data $\{\boldsymbol{X}^v, \forall 1 \le v \le M\}$, the hyperparameters λ_R and λ_C, the loss function $\boldsymbol{\mathcal{F}}(\cdot, \cdot)$, the constraint $\boldsymbol{C}^v$, the learned graph manner for $\boldsymbol{G}_*$.

```
for v=1:M do
    Construct G^v in the loss function F(·,·).
    Initialize U^v by minimizing the loss function F(·,·) under the constraint C^v.
end
while not converged do
    for v=1:M do
        Update G_*^v for the vth view according to the construction manner of the chosen learned graph method.
    end
    for v=1:M do
        Update U^v for the vth view by solving (10).
    end
end
```

Output: Learned representations $\{\boldsymbol{U}^v, 1 \le v \le M\}$.

2) Convergence Analysis: Because we adopt the alternating optimization strategy to solve our proposed framework, it is essential to analyze its convergence.

Theorem 1: The objective function in (8) is bounded. The proposed optimization algorithm monotonically decreases the loss value in each step, which makes the solution converge to a local optimum.

Proof: In most cases of graph embedding loss function in vth view, $\boldsymbol{\mathcal{F}}(\boldsymbol{G}^v, \boldsymbol{U}^v)$ is positive. Thus, it is readily to be satisfied that there must exist one view which can make

$\mathcal{F}_{\min} = \mathcal{F}(\boldsymbol{G}^v, \boldsymbol{U}^v) > 0$ to be smallest among all views. Similarly, we also find that the smooth regularized term $\boldsymbol{\Omega}(\boldsymbol{U}^v)$ must be greater than 0. For the graph consensus terms among views, we could verify that $\mathrm{tr}\left(\boldsymbol{U}^v \boldsymbol{L}^w \boldsymbol{U}^{vT}\right)$ is positive-definite quadratic function if $\boldsymbol{L}^w$ is a positive-definite matrix. Fortunately, this condition is usually established. Similar to the discussion the loss function in each view, there must exist two closest views which could make $\mathcal{C}_{\min} = \mathrm{tr}\left(\boldsymbol{U}^v \boldsymbol{L}^w \boldsymbol{U}^{vT}\right) > 0$ to be smallest among all pairwise views. And because the hyperparameters $\lambda_R > 0$ and $\lambda_C > 0$, it is provable that the objective value in (8) is greater than $M\mathcal{F}_{\min} + M(M-1)\mathcal{C}_{\min}$. Therefore, the objective function in (8) has a lower bound.

For each iteration of optimizing problem (8), we could obtain the learned representations $\{\boldsymbol{U}^v, 1 \le v \le M\}$ by iterative solving (10), which are corresponding to the exact minimum points of (8) for all views, respectively. Under the condition that $\boldsymbol{G}_*^v$ is set to be stationary, the value of the objective function in (10) is nonincreasing in each iteration of **Algorithm 1**. Thus the alternating optimization procedure will monotonically nonincreasing the objective in (8).

Denote the value of loss function in (8) as $\mathcal{H}$, and let $\{\mathcal{H}^t\}_{t=1}^T$ be a sequence generated by the iteration steps in **Algorithm 1**, where T is the length of this sequence. Based on the above analysis, $\{\mathcal{H}^t\}_{t=1}^T$ is a bounded below monotone decreasing sequence. According to the bounded monotone convergence theorem [55] that asserts the convergence of every bounded monotone sequence, the proposed optimization algorithm converges. Accordingly, **Theorem 1** has been proven.

D. Discussion With Other Related Methods

In this section, we give a comprehensive explanation for the proposed GCMF, by discussing the differences and relations between GCMF and other related methods.

Compared with the variants based on CCA, our proposed graph consensus term is not limited by the dimensional equivalent across different views. For the HSIC term in (2), linear kernel is usually used to implement $\boldsymbol{K}_X$ and $\boldsymbol{K}_Y$. Even though this way is convenient to obtain the optimal solution, the optimization for the nonlinear case is not efficient. Besides, Co-reg [19] might meet the similar issue when facing nonlinear cases. Note that, when the graph consensus term focuses on the similarity among samples in other views, HSIC term and the disagreement term in Co-reg could be seen as special cases of the graph consensus term. For example, if $\mathrm{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w) = \boldsymbol{U}^v \boldsymbol{H} \boldsymbol{K}^w \boldsymbol{H} \boldsymbol{U}^{v^T}$, it is equivalent to the definition of HSIC term with linear kernel. Differently, we can flexibly choose the common kernel function as similarity measure for $\boldsymbol{K}^w$, such as Gaussian kernel and graph structure within data, which is more applicable for the nonlinear case.

Compared with those graph structure fusion (GSF)-based works [13], [14], [15], [16] that fuse multiple graphs into one common latent space shared by all views, the proposed GCMF might pay more attention to the complementary efforts between views. Besides, its variants can be scalably to introduce the common embedding in our framework based on the regularization term $\mathrm{Reg}(\boldsymbol{U}, \boldsymbol{G}_*^v)$, where $\boldsymbol{U}$ denotes the common embedding for all views. When explicitly considering the complementary efforts, the regularization term $\mathrm{Reg}(\boldsymbol{U}, \boldsymbol{G}_*^v)$ should be added into (10) to update each view; otherwise, $\sum_{1 \le v \ne w}^{M} \mathrm{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w)$ in (10) should be just substituted with the regularization term $\mathrm{Reg}(\boldsymbol{U}, \boldsymbol{G}_*^v)$. In contrast to above graph-based multiview works, another classical type of graph fusion-based works [8], [50], [54] aim to learn the consensus graph for all views. Even though the proposed GCMF mainly focuses on the embeddings for multiple views, its variants can be also readily extended into such case. For example with GSF [50], first, the shared embedding $\boldsymbol{U}$ is used to approximate the fused affinity matrix; then, the regularization term $\mathrm{Reg}(\boldsymbol{U}, \boldsymbol{G}_*)$ is equal to graph approximation term in GSF; finally, the low-rank constraint is added on the consensus graph $\boldsymbol{G}_*$. In this way, we can implement the transform process from GCMF to GSF.

By comparing the proposed GCMF with its related works, we can summarize the following advantages in terms of exploitation for multiview information and the flexibility of GCMF.

1) For most of existing multiview learning frameworks, the limitation of dimensional equivalent makes it not flexible for the extensions of those works. Differing from those methods, GCMF can flexibly formulate the dimensionality of each view, which eliminates this limitation. More importantly, GCMF can incorporate nonlinear universal cases by exploiting the graph structure information based on learned representations.
2) GCMF is a flexible and scalable multiview framework, which not only can extend most single-view graph embedding methods into the multiview scenario, but also can maintain compatibility with existing GSF methods. Furthermore, to preserve the stability of the multiview framework, it co-regularizes different views through the graph consensus term based on learned graphs, meanwhile preserving the intrinsic property of each view.

IV. Specific Implement

In this section, we choose two graph embedding methods, consisting of LE [56] and LLE [29], to provide a typical implement for our proposed framework, named GCMF-LLE.

A. Construction Process of GCMF-LLE

LLE lies on the manifold structure of the samples space to preserve the relationships among samples. Based on the assumption that each sample and its neighbors lie on or close to a locally linear patch of the manifold, then we obtain the weights matrix $\boldsymbol{S}^v \in \mathbb{R}^{N \times N}$ by minimizing the following reconstruction error:

$$\sum_{i=1}^{N} \left\| \boldsymbol{X}_i^v - \sum_{j \in N\{i\}} \boldsymbol{S}_{ij}^v \boldsymbol{X}_j^v \right\|_2^2 \tag{13}$$

where $N\{i\}$ denotes the neighbors of the ith sample $\boldsymbol{X}_i^v$. By solving the above equation, we could obtain graph structure $\boldsymbol{S}^v$ to reflect the intrinsic properties of the samples space. We expect their characterization of local geometry in the original space to be equally valid for local patches on the

manifold. Each original sample $\boldsymbol{X}_i^v$ is mapped to a new d^v-dimensional coordinate. Additionally, we constrain the learned representations $\boldsymbol{U}_i^v$, $1 \le i \le N$ to have unit covariance. With simple algebraic formulation, the above cost problem can be further transformed as follows:

$$\begin{aligned} &\min_{\boldsymbol{U}^v} \ \operatorname{tr}\left(\boldsymbol{U}^v(\boldsymbol{I}-\boldsymbol{S}^v)^T(\boldsymbol{I}-\boldsymbol{S}^v)\boldsymbol{U}^{vT}\right) \\ &\text{s.t.} \ \ \boldsymbol{U}^v\boldsymbol{U}^{vT} = \boldsymbol{I}. \end{aligned} \tag{14}$$

Hereto, we determine that $\mathcal{F}(\boldsymbol{U}^v)$ and $\boldsymbol{C}^v$ are responding to $\operatorname{tr}(\boldsymbol{U}^v(\boldsymbol{I}-\boldsymbol{S}^v)^T(\boldsymbol{I}-\boldsymbol{S}^v)\boldsymbol{U}^{vT})$ and $\boldsymbol{U}^v\boldsymbol{U}^{vT} = \boldsymbol{I}_N$, respectively.

LE aims at preserving the local neighborhood structure on the data manifold, which constructs the weight matrix that describes the relationships among the samples. Specifically, the similarity matrix $\boldsymbol{G}_*$ is to denote the weight coefficients, which could choose the common kernel function as our similarity measure, such as linear kernel, polynomial kernel, and Gaussian kernel. Combining this with the graph consensus term in (5) between the v view and wth view, we could define $\boldsymbol{L}^w$ as follows:

$$\boldsymbol{L}^w = \boldsymbol{D}^w - \boldsymbol{G}_*^w \tag{15}$$

where $\boldsymbol{D}^w$ denotes a diagonal matrix and $\boldsymbol{D}_{ii}^w = \sum_j \boldsymbol{G}_{*ij}^w$. By rewriting the normalized matrix $\boldsymbol{L}^w$, we could get $\boldsymbol{L}^w = \boldsymbol{I}_N - \boldsymbol{D}^{w-1/2}\boldsymbol{G}_*^w\boldsymbol{D}^{w-1/2}$. Therefore, we can obtain the following graph consensus term between two views:

$$\operatorname{Reg}(\boldsymbol{U}^v, \boldsymbol{G}_*^w) = \operatorname{tr}\left(\boldsymbol{U}^v(\boldsymbol{I}_N - \boldsymbol{D}^{w-1/2}\boldsymbol{G}_*\boldsymbol{D}^{w-1/2})\boldsymbol{U}^{vT}\right). \tag{16}$$

According to the above construction of single-view graph loss function and graph consensus term between views, we have specified each term in objective function in (8) and its constraint terms. In this way, we could extend single-view-based LLE into multiview setting, named multiview LLE (GCMF-LLE). Accordingly, the whole objective function for GCMF-LLE can be formulated as follows:

$$\begin{aligned} &\min \boldsymbol{O}\left(\boldsymbol{U}^1, \boldsymbol{U}^2, \ldots, \boldsymbol{U}^M\right) \\ &\quad = \sum_{v=1}^{M} \operatorname{tr}\left(\boldsymbol{U}^v(\boldsymbol{I}-\boldsymbol{S}^v)^T(\boldsymbol{I}-\boldsymbol{S}^v)\boldsymbol{U}^{vT}\right) + \lambda_R \sum_{v=1}^{M} \boldsymbol{\Omega}(\boldsymbol{U}^v) \\ &\qquad + \lambda_C \sum_{v\neq w} \operatorname{tr}\left(\boldsymbol{U}^v\left(\boldsymbol{I}_N - \boldsymbol{D}^{w-1/2}\boldsymbol{G}_*^w\boldsymbol{D}^{w-1/2}\right)\boldsymbol{U}^{vT}\right) \\ &\text{s.t.} \ \boldsymbol{U}^v\boldsymbol{U}^{vT} = \boldsymbol{I}, \quad 1 \le v \le M. \end{aligned} \tag{17}$$

Because the constraint terms normalize the scale of $\{\boldsymbol{U}^1, \boldsymbol{U}^2, \ldots, \boldsymbol{U}^M\}$, the smooth regularized term $\boldsymbol{\Omega}(\boldsymbol{U}^v)$ could be neglected in the objective function of GCMF-LLE. That is, the above equation could be reduced as follows:

$$\begin{aligned} &\min \boldsymbol{O}\left(\boldsymbol{U}^1, \boldsymbol{U}^2, \ldots, \boldsymbol{U}^M\right) \\ &\quad = \sum_{v=1}^{M} \operatorname{tr}\left(\boldsymbol{U}^v(\boldsymbol{I}-\boldsymbol{S}^v)^T(\boldsymbol{I}-\boldsymbol{S}^v)\boldsymbol{U}^{vT}\right) \\ &\qquad + \lambda_C \sum_{v\neq w} \operatorname{tr}\left(\boldsymbol{U}^v(\boldsymbol{I}_N - \boldsymbol{D}^{w-1/2}\boldsymbol{G}_*^w\boldsymbol{D}^{w-1/2})\boldsymbol{U}^{vT}\right) \\ &\text{s.t.} \ \ \boldsymbol{U}^v\boldsymbol{U}^{vT} = \boldsymbol{I}, \quad 1 \le v \le M. \end{aligned} \tag{18}$$

B. Optimization

Referring to the optimization procedure for GCMLF, (18) could be approximately solved. When solving the vth view, with all views fixed but $\boldsymbol{U}^v$, we get the following optimization for the vth view:

$$\begin{aligned} \min \boldsymbol{O}(\boldsymbol{U}^v) &= \operatorname{tr}\left(\boldsymbol{U}^v(\boldsymbol{I}-\boldsymbol{S}^v)^T(\boldsymbol{I}-\boldsymbol{S}^v)\boldsymbol{U}^{vT}\right) \\ &\quad + \lambda_C \sum_{1\le v\neq w}^{M} \operatorname{tr}\left(\boldsymbol{U}^v(\boldsymbol{I}_N - \boldsymbol{D}^{w-1/2}\boldsymbol{G}_*^w\boldsymbol{D}^{w-1/2})\boldsymbol{U}^{vT}\right) \\ \text{s.t.} \ \ \boldsymbol{U}^v\boldsymbol{U}^{vT} &= \boldsymbol{I}. \end{aligned} \tag{19}$$

Due to the attributes of the matrix trace, the above equation is equivalent to the following optimization problem:

$$\begin{aligned} \min \boldsymbol{O}(\boldsymbol{U}^v) = \operatorname{tr}\Bigg(&\boldsymbol{U}^v((\boldsymbol{I}-\boldsymbol{S}^v)^T(\boldsymbol{I}-\boldsymbol{S}^v) \\ &+ \lambda_C \sum_{1\le v\neq w}^{M} (\boldsymbol{I}_N - \boldsymbol{D}^{w-1/2}\boldsymbol{G}_*^w\boldsymbol{D}^{w-1/2})\boldsymbol{U}^{vT}\Bigg) \\ \text{s.t.} \ \ \boldsymbol{U}^v\boldsymbol{U}^{vT} &= \boldsymbol{I}. \end{aligned} \tag{20}$$

Under the constraint condition $\boldsymbol{U}^v\boldsymbol{U}^{vT} = \boldsymbol{I}$, the above equation could be efficiently solved by eigenvalue decomposition. In this way, we could solve all the variables $\{\boldsymbol{U}^v, \boldsymbol{G}_*^v, 1 \le v \le M\}$ iteratively.

According to the convergence analysis for our framework in Section III-C, it could be easily verified that the optimization procedure for GCMF-LLE will be converged within limited iteration steps. We also use many experiments to verify the convergence property of the proposed method. Fig. 2 shows the relation between the objective values and iterations. As shown in Fig. 2, we can see that with the iterations increase, the objective function value of the proposed method decreases fast and reaches a stable point after a few iterations, while the classification accuracy increases dramatically during the first small number of iterations and then reaches the stable high level for these four benchmark databases. For example, for the Holidays dataset, the proposed method reaches the stable point in terms of classification accuracy within about fifteen iterations. Both theoretical proof and experiments demonstrate that the proposed method can obtain the local optimum quickly and has good convergence property.

C. Time Complexity Analysis

The computational cost for GCMF-LLE mainly is composed of two parts. One is the construction for the variables $\{\boldsymbol{S}^v, i \le v \le M\}$ and the initialization for the variables and $\{\boldsymbol{U}^v, i \le v \le M\}$, which solves $\boldsymbol{S}^v$ and $\boldsymbol{U}^v$ according to (13) and (14). The other is to iteratively update $\boldsymbol{K}^v$ and $\boldsymbol{U}^v$, which needs to perform the computation of similarity matrix and eigenvalue decomposition in each iteration, respectively. Therefore, the time complexity for GCMF-LLE is about $\mathrm{O}(T \times M \times N^3)$, where T is the iteration times of the alternating optimization procedure. Note that, based on the convergence of the optimization procedure of GCMF-LLE, the iteration times T will be a limited number. Therefore, its time complexity is linear

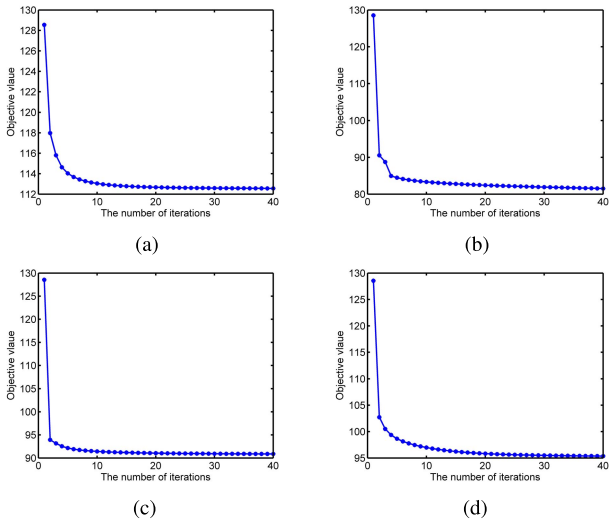


Fig. 2. Convergence validations on four datasets. (a) Yale dataset. (b) Holidays dataset. (c) ORL dataset. (d) Corel-1K dataset.

with respect to LLE, which can imply that the optimization for GCML is efficient.

D. Discussion

LLE and LE are two graph embedding methods based on self-representation and geometric structure styles, respectively, in which LLE is used to construct the graph learning loss term and LE is used to regularize the dependence between two views in (8). Note that, LLE is based on manifold space reconstruction, which aims to preserve reconstruction relationships among samples. Therefore, when LE is utilized to construct the graph learning loss term, we also consider that LLE is used to construct the graph consensus term between two views by (7). To facilitate the solution, we choose the former to specify the graph learning loss term in (8) in this article.

V. Experiments

In this section, we introduce the details of several experiments on document classification, face recognition, and image retrieval, to verify the effectiveness of our proposed framework.

A. Datasets and Compared Methods

In our experiments, six datasets are used to validate the superior performance of our framework, including document datasets (3Source[1] and Cora[2]), face datasets(ORL[3] and Yale[4]), and image datasets(Corel-1K[5] and Holidays[6]). Two document datasets are two benchmark multiview datasets. For the face and image datasets, we utilize different descriptors to extract their corresponding multiview features, in which some samples in these datasets are shown in Fig. 3. The detailed information of these datasets is summarized as follows.

1) *3Source* consists of three well-known news organizations: BBC, Reuters, and Guardian, where each news is manually annotated with one of six labels. Because each news source can be used as one view, we choose these news sources as a multiview benchmark dataset.
2) *Cora* contains 2708 scientific publications of seven categories, where each publication document could be described by content and citation. Thus, Cora could be considered as a two-view benchmark dataset.
3) *ORL* is collected from 40 distinct subjects, where ten different images are gathered for each subject. For each person, the images are taken at different times, varying the lighting, facial expressions, and facial details.
4) *Yale* is composed of 165 faces from 15 peoples, which has been widely used in face recognition. Each person has eleven images, with different facial expressions and facial details.
5) *Corel-1K* manually collects 1000 images corresponding to ten categories, such as human beings, buildings, landscapes, buses, dragons, elephants, horses, flowers, mountains, and foods. And there are one hundred images in each category.
6) *Holidays* consists of 1491 images corresponding to 500 categories, which are mainly captured for sceneries.

Even though text and images adopted in experiments don't explicitly contains the graph structure information, there exists the relationship among samples in the above datasets, such as similarity and reconstruction relationships. Based on the similarity or reconstruction relationship among samples, the proposed GCMF can build the graph-structure data for all views to exploit their intrinsic information among multiple views, where each sample (text or image) can be seen as one node in the graph.

To demonstrate the superior performance of our framework, we compare GCMF-LLE with the following methods, where the first two are single-view methods with the most informative view, and the others are multiview learning methods.

1) *BLE* is Laplacian eigenmaps (LE) [56] with the most informative view, i.e., one that achieves the best performance with LE.
2) *BLLE* is LLE [29] with the most informative view, similar to BLE.
3) *GFSC* [53] is a multiview spectral embedding based on multigraph fusion to approximate the original graph of individual view.
4) *GMC* [8] is a multiview graph-based method to learn the common graph shared by all views.
5) *GMA* [21] is a general multiview learning framework, solving the joint and relaxed problem of the form of QCQP.
6) *MDcR* [39] is a multiview dimensionality reduction method, which explores the correlations of different views based on HSIC term.

[1] http://mlg.ucd.ie/datasets/3sources.html
[2] http://lig-membres.imag.fr/grimal/data.html
[3] http://www.U.K..research.att.com/facedatabase.html
[4] http://cvc.yale.edu/projects/yalefaces/yalefaces.html
[5] https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval
[6] http://lear.inrialpes.fr/jegou/data.php

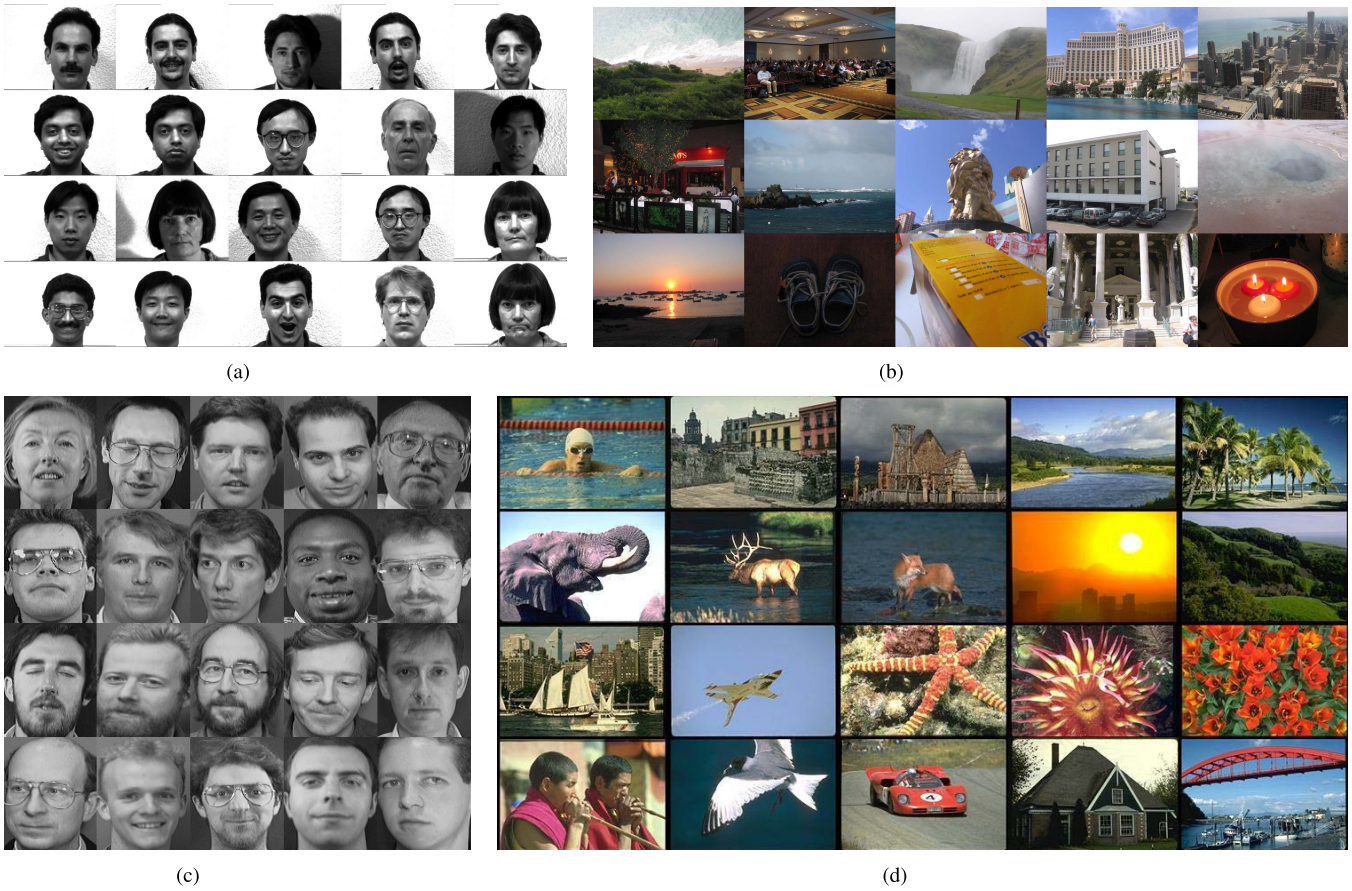


Fig. 3. Examples images in datasets. (a) Some examples in Yale dataset. (b) Some examples in Holidays dataset. (c) Some examples in ORL dataset. (d) Some examples in Corel-1K dataset.

7) *AMGL* [14] is an auto-weighted multiple graph learning method, which could allocate ideal weight for each view automatically.

B. Document Classification

In this section, we evaluate the experimental results of the document classification tasks on 3Source and Cora datasets. For these two datasets, we randomly select 50% of the samples as training samples and the remaining 50% of the dataset as testing samples every time. All the methods are conducted to project all samples to the same dimensionality. Specifically, the dimensions of the embedding obtained by all methods all maintain 20 and 30 dimensions. We adopt 1NN as the classifier to classify the testing ones. After conducting this experiment 30 times with different random training samples and testing samples, we calculate the mean classification accuracy (MEAN) and max classification accuracy (MAX) on 3Source and Cora datasets as the evaluation index for all methods. Then, we can summarize the evaluation indexes of MEAN and MAX results in Tables II and III.

Through the experimental results of Tables II and III, it is clear that the proposed GCMF-LLE is significantly superior to its counterparts in most situations. Besides, the performance of the GCMF-LLE is more stable than other compared methods. For example, GMC can obtain promising results on 3Source dataset while the performance degrades sharply on the Cora dataset; in contrast to GMC, GMA can obtain the superior performance on the Cora dataset but get the poor results on the 3Source dataset.

TABLE II
CLASSIFICATION ACCURACY ON 3SOURCE DATASET

Methods	Dims=20		Dims=30	
	MEAN(%)	MAX(%)	MEAN(%)	MAX(%)
BLE	66.47	74.11	59.72	69.41
BLLE	66.50	76.71	66.78	75.94
GFSC	76.01	84.31	80.30	88.41
GMC	80.39	88.23	80.84	90.19
GMA	53.88	76.45	54.37	73.56
MDcR	81.25	87.05	78.50	85.88
AMGL	49.92	57.64	48.15	56.47
GCMF-LLE	**82.64**	**89.41**	**81.25**	**90.93**

TABLE III
CLASSIFICATION ACCURACY ON CORA DATASET

Methods	Dims=20		Dims=30	
	MEAN(%)	MAX(%)	MEAN(%)	MAX(%)
BLE	58.98	60.85	61.05	63.44
BLLE	59.84	63.61	60.86	65.31
GFSC	38.39	42.18	39.04	42.18
GMC	42.06	44.64	42.06	44.77
GMA	71.11	72.35	71.52	72.05
MDcR	55.73	57.45	57.19	59.01
AMGL	63.71	65.73	66.90	69.57
GCMF-LLE	**73.7**	**75.23**	**73.45**	**75.84**

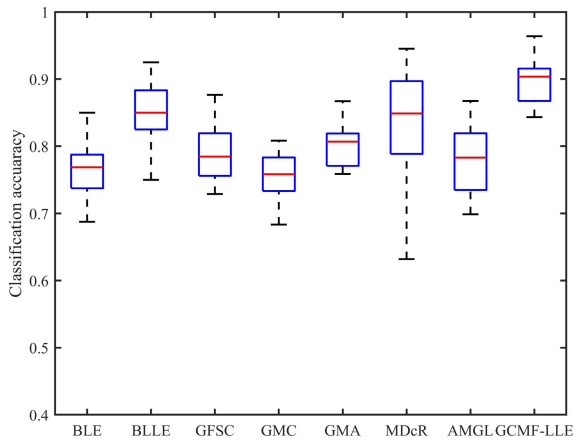


Fig. 4. Face recognition accuracy on Yale dataset.

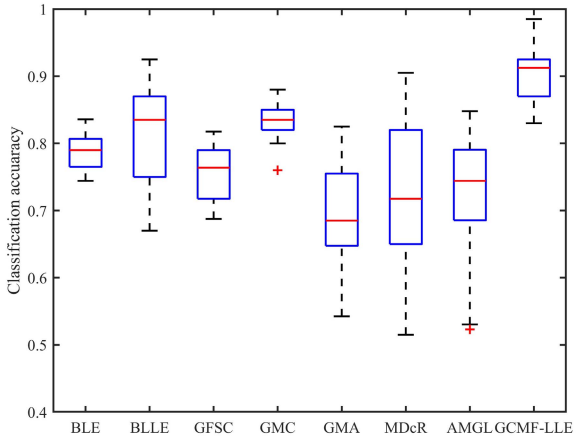


Fig. 5. Face recognition accuracy on ORL dataset.

C. Face Recognition

In this section, we evaluate the experimental results of the face recognition tasks on Yale and ORL datasets. For these two datasets, we first extract their multiview features by EDH [57], LBP [58], and Gist [1]. Then, all the methods are conducted to project all samples to the same dimensionality and the 1NN classifier is adopted to calculate the recognition results, where the dimension of the embedding all maintains 30 dimensions. Note that we randomly select 50% of the samples as training samples and the remaining 50% of the samples as testing samples every time and run all methods 30 times with different random training samples. Because the task of face recognition mainly cares about recognition accuracy, we choose recognition accuracy as the evaluation index in this part. The boxplot figures of accuracy values of all methods on Yale and ORL datasets are shown in Figs. 4 and 5.

Through the experiment results of the above two experiments in Figs. 4 and 5, the multiple view performances are usually better than the independent view. This demonstrates that multiple views can improve the performance of face recognition. Among these multiview methods, we can find that GCMF-LLE outperforms its comparing methods in most situations, which shows the superiority of the proposed framework.

TABLE IV
IMAGE RETRIEVAL ACCURACY ON HOLIDAYS DATASET

Methods	Precision (%)	Recall (%)	MAP (%)	F_1-Measure
BLE	72.92	56.16	86.46	31.73
BLLE	59.84	63.61	80.86	30.73
GFSC	65.82	50.43	79.76	28.55
GMC	69.16	66.18	78.54	33.18
GMA	65.22	50.05	78.32	28.32
MDcR	78.49	60.72	88.52	34.24
AMGL	68.09	51.92	84.01	29.46
GCMF-LLE	**79.13**	**61.14**	**89.56**	**34.49**

D. Image Retrieval

In this section, we conduct two experiments on Holidays and Corel-1K datasets for image retrieval. For these two datasets, we both employ three image descriptors of MSD [59], Gist [1], and HOC [60] to extract multiview features for all images. All the methods are conducted to project all samples to the same dimensionality. In this part, the dimensions of the embedding obtained by all methods maintain 30 dimensions. Besides, l_1 distance is utilized to measure similarities between samples. At the aspect of the validation index, we choose several common indexes, including average precision rate (Precision), average recall rate (Recall), mean average precision (MAP), and F_1-Measure, to validate the performances for image retrieval. Actually, high Precision and Recall are required and F_1-Measure is put forward as the overall performance measurement. Then, we conducted this experiment on these two datasets repeatedly for 20 times. For Holidays dataset, we summarize these experiment results, including Precision, Recall, MAP, and F_1-Measure, on top 2 retrieval results in Table IV. For Corel-1K dataset, we randomly select ten images as query ones for each category. Afterward, the relation curves on validation indexes are drawn in Fig. 6.

Through these experimental results in Table IV and Fig. 6, it can be readily found that our proposed GCMF-LLE achieves better performance than the other compared methods in most situations in the field of image retrieval. The proposed GCMF-LLE could integrate compatible and complementary information from multiple views and obtain a better embedding from these views. Therefore, the results in Table IV and Fig. 6 could show that our framework can achieve good performance in the field of face recognition. Note that the performance of BLE is bad because of its unreasonable way to deal with multiview features.

E. Sensitivity Analysis

To fully validate the effectiveness of GCMF-LLE, this subsection mainly analyzes the influence on the performance of the parameter λ_C introduced in GCMF-LLE. As is shown in Fig. 7, which summarizes the classify accuracy values of 3Source, Cora, ORL, and Yale datasets, where the dimensionality of low-dimensional embedding is 30. It is easy to find that the oscillation of accuracy becomes very stable, which

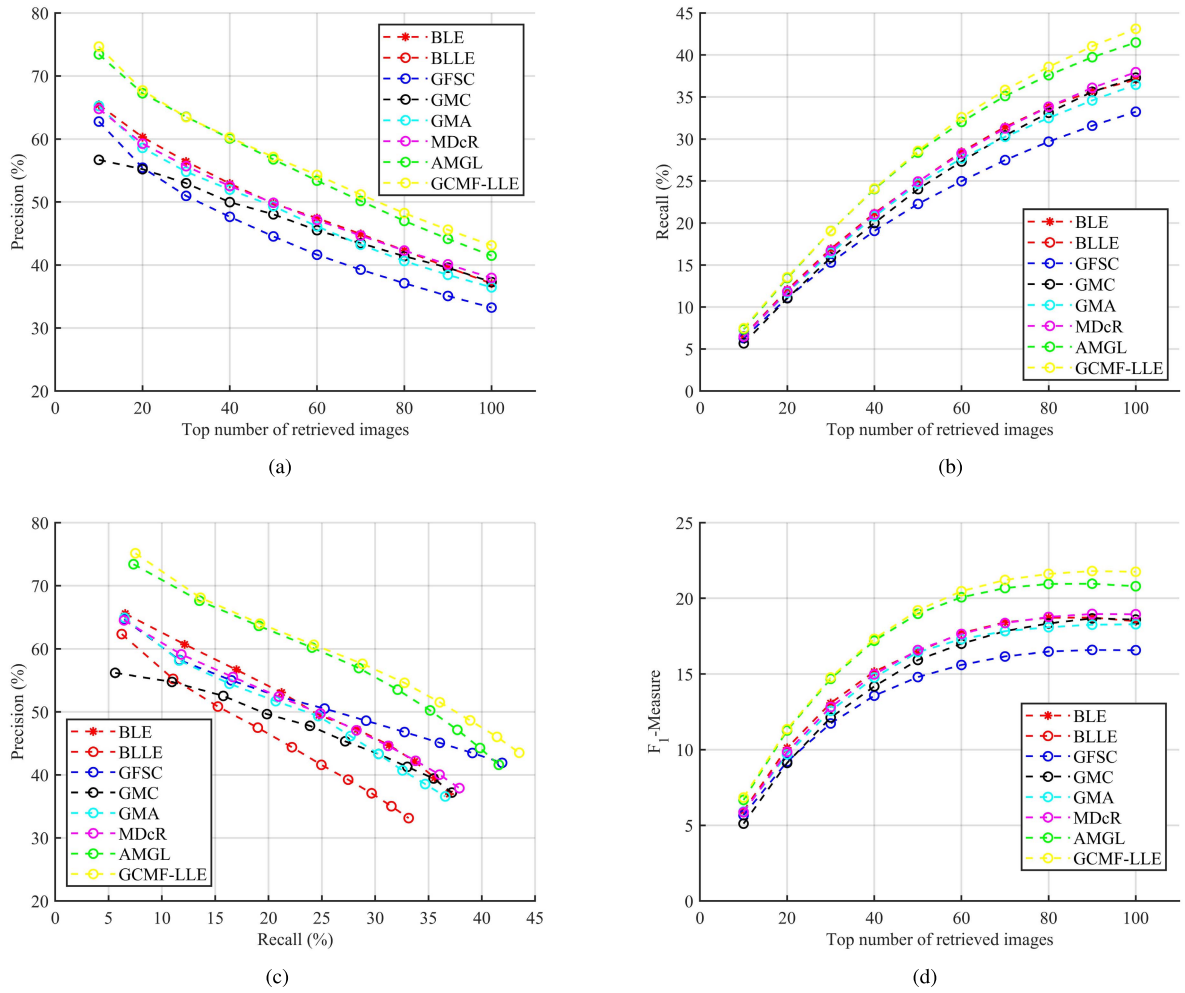


Fig. 6. Curves of precision, recall, PR, and F_1-Measure on Corel-1K dataset. (a) Precision. (b) Recall. (c) PR-Curve. (d) F_1-Measure.

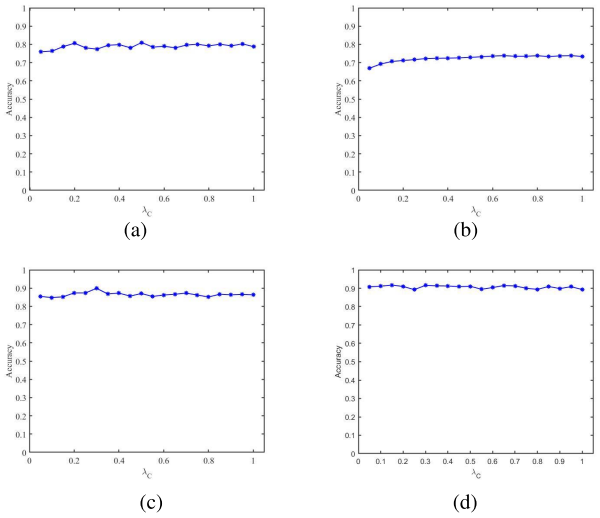


Fig. 7. Sensitivity analysis on four datasets. (a) 3Source dataset. (b) Cora dataset. (c) Yale dataset. (d) ORL dataset.

indicates that the performance is not so sensitive to those hyperparameters. More importantly, there exists a wide range for each hyper-parameter in which relatively stable and good results can be readily obtained.

F. Stability Analysis

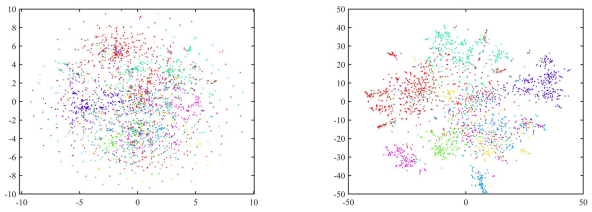
To validate the model stability of GCMF-LLE, we conduct the cross-validation experiments under different settings. To be specific, we run the twofold, threefold, fivefold, and tenfold cross-validation experiments on the 3Source, Cora, ORL, and Yale datasets, respectively. For the example of fivefold cross-validation, onefold and the other fourfold are used for testing data and training data, respectively, thus the validation process is repeated five times, and the average accuracy over these five runs is used as the final result. And the dimensionality of low-dimensional embedding is 30. We summarize the average accuracy values of different cross-validation settings on the four datasets in Table V. Through the results in Table V, we can find that the variation of the performance of cross-validation under different settings is relatively stable. That is, the proposed GCML is a stable multiview learning framework.

G. Visualization of GCMF-LLE

To visualize the sample distribution learned by GCMF-LLE, we first adopt t-SNE [61] to project original data and learned features into the 2-D subspace, and then visualize their distributions. This experiment is conducted on the Cora

TABLE V
CROSS-VALIDATION RESULTS (%) ON FOUR DATASETS

Setting	Cora	3Source	Yale	ORL
2-fold	73.19	79.84	87.06	89.14
3-fold	73.82	79.27	87.17	90.79
5-fold	74.41	81.14	88.68	92.53
10-fold	74.96	82.23	90.43	92.55



(a) (b)

Fig. 8. Visualization of GCMF-LLE on Cora dataset. (a) Original data. (b) GCMF-LLE.

dataset, and we visualize the learned features in the first view, shown in Fig. 8. Obviously, the distributions of original data are disordered. After GCMF-LLE is conducted, the samples can be readily separated into several clusters, which can validate the effectiveness of GCMF-LLE.

H. Discussion

For the experiment results in Tables II and III on text classification, we can find that GCMF-LLE outperforms other comparing methods in most situations. Similarly, our proposed GCMF-LLE also obtain promising performance in face recognition tasks through the evaluations in Figs. 4 and 5. As shown in Table IV and Fig. 6, our method could also be utilized to execute the image retrieval task. From the above evaluations, it is readily seen that the representations obtained by our method could be more effective and suitable for multiview features.

According to the above experimental results, we can drive the following findings. Compared with BLLE and BLE, GCMF-LLE could achieve significantly better performance by integrating complementary information among different views meanwhile preserving its intrinsic characteristic in each view. Compared with other multiview methods, GCMF-LLE can obtain more robust and efficient performance due to flexibility and stability of GCMF. Note that the experimental results of our proposed GCMF-LLE on six datasets are without fine-tuning for the views' weights, and usage of fine-tuning (self-weighting or grid searching strategy) might further improve its performance. Besides, we empirically find that GCMF-LLE could converge within limited iterations in most experiments.

Notably, graph convolution network (GCN) [62], [63] has gained extensive attention from researchers, which is also considered as graph-based work. Different from the graph-based work investigated in this article, GCN is built on the explicit graph structure and label information. GCMF-LLE is an unsupervised multiview method, and those datasets in the experiments cannot provide explicit graphs besides Cora. Even though GCN is not suitable to be utilized as a comparing method in this article, we aim to combine our proposed GCMF with GCN to solve the graph learning problems under the multiview scenario.

VI. CONCLUSION

In this article, we propose a unified and scalable multiview learning framework, named GCMF, which aims at leveraging most existing graph embedding works into one formula via introducing the graph consensus term. GCMF encourages all views to learn with each other according to the complementarity among views and explores the learned graph structure in each view independently to preserve the diversity property among all views. Based on the sufficient theoretical analysis, we show that GCMF is a more robust and flexible multiview learning framework than those existing multiview methods. Correspondingly, an algorithm based on alternating direction strategy is proposed to solve GCMF. To further facilitate the related research and the understanding of GCMF, we provide one typical implementation of the multiview extension for LLE, called GCMF-LLE. Extensive experimental results demonstrate that the proposed GCMF-LLE can effectively explore the diversity information and underlying complementary information of the given multiview data, and outperforms its compared methods. With the rapid development of graph neural networks [64], [65], [66], how to combine our proposed GCMF with GCN to solve the multiview problems with graph information is very meaningful yet full of challenges, and we will consider it in our future work.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their insightful comments and suggestions to significantly improve the quality of this article.

REFERENCES

[1] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg, and C. Schmid, "Evaluation of GIST descriptors for web-scale image search," in *Proc. ACM Int. Conf. Image Video Retr. (CIVR)*, 2009, pp. 1–8.
[2] G. Bisson and C. Grimal, "Co-clustering of multi-view datasets: A parallelizable approach," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 828–833.
[3] G. Chao, S. Sun, and J. Bi, "A survey on multiview clustering," *IEEE Trans. Artif. Intell.*, vol. 2, no. 2, pp. 146–168, Apr. 2021.
[4] 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.
[5] M. Kan, S. Shan, and X. Chen, "Multi-view deep network for cross-view classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4847–4855.
[6] C. Zhang, J. Cheng, and Q. Tian, "Multi-view image classification with visual, semantic and view consistency," *IEEE Trans. Image Process.*, vol. 29, pp. 617–627, 2019.
[7] Q. Yin, S. Wu, and L. Wang, "Multiview clustering via unified and view-specific embeddings learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5541–5553, Nov. 2018.
[8] 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, May 2019.
[9] L. Feng, X. Meng, and H. Wang, "Multi-view locality low-rank embedding for dimension reduction," *Knowl.-Based Syst.*, vol. 191, Mar. 2020, Art. no. 105172.

[10] X. Meng, L. Feng, and H. Wang, "Multi-view low-rank preserving embedding: A novel method for multi-view representation," *Eng. Appl. Artif. Intell.*, vol. 99, Mar. 2021, Art. no. 104140.

[11] H. Wang, J. Peng, D. Chen, G. Jiang, T. Zhao, and X. Fu, "Attribute-guided feature learning network for vehicle reidentification," *IEEE Multimedia*, vol. 27, no. 4, pp. 112–121, Oct. 2020.

[12] H. Wang, J. Peng, Y. Zhao, and X. Fu, "Multi-path deep CNNs for fine-grained car recognition," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 10484–10493, Oct. 2020.

[13] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 6, pp. 1438–1446, Dec. 2010.

[14] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, Mar. 2018.

[15] S. Huang, Z. Kang, and Z. Xu, "Self-weighted multi-view clustering with soft capped norm," *Knowl.-Based Syst.*, vol. 158, no. 15, pp. 1–8, Oct. 2018.

[16] L. Tian, F. Nie, and X. Li, "A unified weight learning paradigm for multi-view learning," in *Proc. 22nd Int. Conf. Artif. Intell. Statist.*, vol. 89, 2019, pp. 2790–2800.

[17] W. Wang and Z. H. Zhou, "A new analysis of co-training," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 1–8.

[18] A. Kumar and H. Daumé, "A co-training approach for multi-view spectral clustering," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 393–400.

[19] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1413–1421.

[20] H. Wang *et al.*, "Kernelized multiview subspace analysis by self-weighted learning," *IEEE Trans. Multimedia*, vol. 23, pp. 3828–3840, 2021.

[21] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2160–2167.

[22] G. Cao, A. Iosifidis, K. Chen, and M. Gabbouj, "Generalized multi-view embedding for visual recognition and cross-modal retrieval," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2542–2555, Sep. 2018.

[23] X. Meng, H. Wang, and L. Feng, "The similarity-consensus regularized multi-view learning for dimension reduction," *Knowl.-Based Syst.*, vol. 199, Jul. 2020, Art. no. 105835.

[24] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[25] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[26] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1473–1480.

[27] L. Torresani and K.-C. Lee, "Large margin component analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1385–1392.

[28] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Proc. Neural Netw. Signal Process., IEEE Signal Process. Soc. Workshop*, Aug. 1999, pp. 41–48.

[29] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[30] D. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004.

[31] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, Jan. 2002.

[32] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Proc. Conf. Data Mining Data Warehouses*, 2010, pp. 1–4.

[33] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 188–194, Jan. 2016.

[34] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

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

[36] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring statistical dependence with Hilbert–Schmidt norms," in *Proc. Int. Conf. Algorithmic Learn. Theory*. Berlin, Germany: Springer, 2005, pp. 63–77.

[37] D. Niu, J. G. Dy, and A. M. I. Jordan, "Iterative discovery of multiple AlternativeClustering views," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1340–1353, Jul. 2014.

[38] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 586–594.

[39] C. Zhang, H. Fu, Q. Hu, P. Zhu, and X. Cao, "Flexible multi-view dimensionality co-reduction," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 648–659, Feb. 2017.

[40] 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.

[41] T. Zhou, C. Zhang, C. Gong, H. Bhaskar, and J. Yang, "Multiview latent space learning with feature redundancy minimization," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1655–1668, Apr. 2020.

[42] J. Li, M. Li, G. Lu, B. Zhang, H. Yin, and D. Zhang, "Similarity and diversity induced paired projection for cross-modal retrieval," *Inf. Sci.*, vol. 539, pp. 215–228, Oct. 2020.

[43] H. Wang, G. Jiang, J. Peng, and X. Fu, "MSAV: An unified framework for multi-view subspace analysis with view consistence," in *Proc. Int. Conf. Multimedia Retr.*, Aug. 2021, pp. 653–659.

[44] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[45] K. Tang, R. Liu, Z. Su, and J. Zhang, "Structure-constrained low-rank representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2167–2179, Dec. 2014.

[46] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Mar. 2013.

[47] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering," in *Proc. Int. Conf. Artif. Intell.*, 2015, pp. 1–7.

[48] G. Ma *et al.*, "Multi-view clustering with graph embedding for connectome analysis," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 127–136.

[49] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple Kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.

[50] K. Zhan, C. Niu, C. Chen, F. Nie, C. Zhang, and Y. Yang, "Graph structure fusion for multiview clustering," *IEEE Trans. Know. Data Eng.*, vol. 31, no. 10, pp. 1984–1993, Oct. 2019.

[51] H. Wang, Y. Yang, B. Liu, and H. Fujita, "A study of graph-based system for multi-view clustering," *Knowl.-Based Syst.*, vol. 163, pp. 1009–1019, Jan. 2019.

[52] R. Wang, F. Nie, Z. Wang, H. Hu, and X. Li, "Parameter-free weighted multi-view projected clustering with structured graph learning," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 10, pp. 2014–2025, Oct. 2020.

[53] Z. Kang *et al.*, "Multi-graph fusion for multi-view spectral clustering," *Knowl.-Based Syst.*, vol. 189, Feb. 2020, Art. no. 105102.

[54] B.-Y. Liu, L. Huang, C.-D. Wang, J.-H. Lai, and P. S. Yu, "Multiview clustering via proximity learning in latent representation space," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 25, 2021, doi: 10.1109/TNNLS.2021.3104846.

[55] W. Rudin *et al.*, *Principles of Mathematical Analysis*, vol. 3. New York, NY, USA: McGraw-Hill, 1964.

[56] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[57] X. Gao, B. Xiao, D. Tao, and X. Li, "Image categorization: Graph edit direction histogram," *Pattern Recognit.*, vol. 41, no. 10, pp. 3179–3191, Oct. 2008.

[58] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[59] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor," *Pattern Recognit.*, vol. 44, no. 9, pp. 2123–2133, 2011.

[60] L. Yu, L. Feng, C. Chen, T. Qiu, L. Li, and J. Wu, "A novel multi-feature representation of images for heterogeneous IoTs," *IEEE Access*, vol. 4, pp. 6204–6215, 2016.

[61] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
[62] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
[63] Y. Zhu, W. Xu, J. Zhang, Q. Liu, S. Wu, and L. Wang, "Deep graph structure learning for robust representations: A survey," 2021, *arXiv:2103.03036*.
[64] Q. Cui, S. Wu, Q. Liu, W. Zhong, and L. Wang, "MV-RNN: A multi-view recurrent neural network for sequential recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 2, pp. 317–331, Feb. 2020.
[65] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.
[66] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Deep graph contrastive representation learning," in *Proc. ICML Workshop Graph Represent. Learn. Beyond*, Jun. 2020, pp. 1–17.

Xiangzhu Meng received the B.S. degree from Anhui University, Hefei, China, in 2015, and the Ph.D. degree in computer science and technology from the Dalian University of Technology, Dalian, China, in 2021.

He is currently a Post-Doctoral Researcher with the Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He regularly publishes articles in prestigious journals, including *Knowledge-Based System* (KBS), *Engineering Applications of Artificial Intelligence* (EAAI), and *Neurocomputing*. His research interests include multiview learning and deep learning.

Lin Feng received the B.S. and M.S. degrees in internal combustion engine and the Ph.D. degree in mechanical design and theory from the Dalian University of Technology, Dalian, China, in 1992, 1995, and 2004, respectively.

He is currently a Professor and a Doctoral Supervisor with the School of Innovation Experiment, Dalian University of Technology. His research interests include intelligent image processing, data mining, and embedded systems.

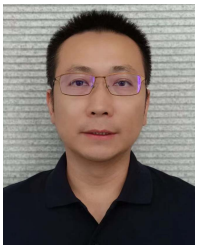
Chonghui Guo received the B.S. degree in mathematics from Liaoning University, Shenyang, China, in 1995, and the M.S. degree in operational research and control theory and the Ph.D. degree with the Institute of Systems Engineering, Dalian University of Technology, Dalian, China, in 1999 and 2002, respectively.

He was a Post-Doctoral Research Fellow with the Department of Computer Science, Tsinghua University, Beijing, China. He is currently a Professor with the Institute of Systems Engineering, Dalian University of Technology. His research interests include data mining and knowledge discovery.

Huibing Wang received the Ph.D. degree from the School of Computer Science and Technology, Dalian University of Technology, Dalian, China, in 2018.

From 2016 to 2017, he was a Visiting Scholar with the University of Adelaide, Adelaide, Australia. He is currently an Associate Professor with Dalian Maritime University, Dalian. He has authored and coauthored more than 50 papers in some famous journals or conferences, including the International Joint Conference on Artificial Intelligence (IJCAI), the IEEE TRANSACTIONS ON MULTIMEDIA (TMM), the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (TITS), the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (TVT), the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS (TSMC), the IEEE MULTIMEDIA (MM), the European Conference on Computer Vision (ECCV), the International Conference on Multimedia Retrieval (ICMR), and the International Conference on Multimedia and Expo (ICME). His research interests include computer vision and machine learning.

Dr. Wang serves as a Reviewer for the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (TKDE), the *ACM Transactions on Information Systems* (ACM TOIS), the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS), the IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS (TCDS), the *ACM Transactions on Multimedia Computing, Communications, and Applications* (ACM TOMM), and *Information Fusion*.



Shu Wu (Senior Member, IEEE) received the B.S. degree from Hunan University, Changsha, China, in 2004, the M.S. degree from Xiamen University, Xiamen, China, in 2007, and the Ph.D. degree from the Department of Computer Science, University of Sherbrooke, Sherbrooke, QC, Canada, in 2012, all in computer science.

He is currently an Associate Professor with the Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China. He has published more than 70 papers in the areas of data mining and information retrieval in international journals and conferences, such as the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (TKDE), the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS (THMS), the Association for the Advancement of Artificial Intelligence (AAAI), the International Conference on Data Mining (ICDM), the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), and the Conference on Information and Knowledge Management (CIKM). His research interests include data mining and information retrieval.