SURVEY PAPER

Representation Learning in Multi-view Clustering: A Literature Review

Man-Sheng Chen[1,2] · Jia-Qi Lin[3] · Xiang-Long Li[1,2] · Bao-Yu Liu[1,2] · Chang-Dong Wang[1,2] · Dong Huang[4] · Jian-Huang Lai[1]

Abstract

Multi-view clustering (MVC) has attracted more and more attention in the recent few years by making full use of complementary and consensus information between multiple views to cluster objects into different partitions. Although there have been two existing works for MVC survey, neither of them jointly takes the recent popular deep learning-based methods into consideration. Therefore, in this paper, we conduct a comprehensive survey of MVC from the perspective of representation learning. It covers a quantity of multi-view clustering methods including the deep learning-based models, providing a novel taxonomy of the MVC algorithms. Furthermore, the representation learning-based MVC methods can be mainly divided into two categories, i.e., shallow representation learning-based MVC and deep representation learning-based MVC, where the deep learning-based models are capable of handling more complex data structure as well as showing better expression. In the shallow category, according to the means of representation learning, we further split it into two groups, i.e., multi-view graph clustering and multi-view subspace clustering. To be more comprehensive, basic research materials of MVC are provided for readers, containing introductions of the commonly used multi-view datasets with the download link and the open source code library. In the end, some open problems are pointed out for further investigation and development.

Keywords Multi-view clustering · Representation learning · Shallow model · Deep model · Data mining

✉ Chang-Dong Wang
changdongwang@hotmail.com

Man-Sheng Chen
chenmsh27@mail2.sysu.edu.cn

Jia-Qi Lin
linjq56@mail2.sysu.edu.cn

Xiang-Long Li
lixlong23@mail2.sysu.edu.cn

Bao-Yu Liu
baoyu.l@hotmail.com

Dong Huang
huangdonghere@gmail.com

Jian-Huang Lai
stsljh@mail.sysu.edu.cn

[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China

[2] Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Guangzhou, China

[3] School of Mathematics (Zhuhai), Sun Yat-sen University, Zhuhai, China

[4] College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China

1 Introduction

Multi-view data obtained from different feature extractors or collected from multiple sources are universal in many real-world applications. For instance, a webpage is often represented by multi-view features based on text, image and video. We can describe a person with face, fingerprint as well as iris. News can be reported by multilingual forms, such as English, Chinese and Italian. Multiple views or feature subsets comprehensively exhibit different angles or aspects of such data. For unsupervised learning, it is crucial to effectively make full use of multiple views or the underlying structures without the guided ground-truth, motivating the development of MVC. In particular, MVC aims to explore and exploit the complementary information from multi-view feature observations to search for a consensus and more robust cluster partition than single-view clustering.

In the past few years, many researchers have made efforts to integrate useful information from different views to improve the clustering performance. Note that there are two important principles for the success of MVC [1], i.e., the consensus principle and the complementary principle.
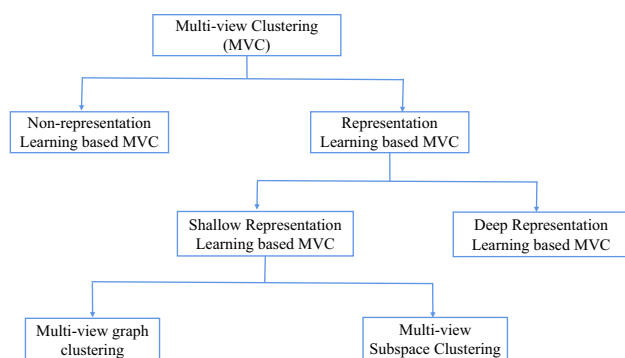
Fig. 1 Taxonomy of representation learning in MVC approaches

Table 1 Difference analysis on two existing MVC surveys and ours. ✓ implies that the corresponding survey includes the listed important part. Otherwise, there would be blank

Materials	Shallow MVC	Deep MVC	Research materials	Open problems
Yang et al. [2]	✓			
Chao et al. [3]	✓			✓
Ours	✓	✓	✓	✓

Specifically, the consensus principle aims to maximize the agreements among different views, and the complementary one shows that each view of the data consists of some information the other views do not have. Among traditional MVC methods, a naive strategy for integrating multiple views is to concatenate the multi-view features into a new feature space, on which single view clustering algorithm (e.g., spectral clustering) would be performed to achieve the clustering performance. Generally, such a naive concatenation pays no attention to the complementary information between multi-view data. To consider more deeper correlations between pairwise views, many advanced learning algorithms for MVC are designed, in which the existing MVC methods are mainly divided into two categories, i.e., non-representation learning-based MVC and representation learning-based MVC. In this paper, we conduct a comprehensive survey of MVC methods from the perspective of representation learning, where a certain inner layer is included from input to output for representation learning. In representation learning-based MVC family, there are mainly two kinds of models to integrate multiple views, i.e., shallow representation learning-based MVC and deep representation learning-based MVC. The emerging deep representation learning-based MVC is capable of handling more complex data structures via deep network framework. For clarity, we summarize the complete taxonomy in Fig. 1. Due to the different strategies of representation learning, we further split those shallow representation learning-based MVC methods into the following two categories, i.e., multi-view graph clustering and multi-view subspace clustering. In particular, multi-view graph clustering attempts to record the relations between different samples with some similarity or distance metrics (e.g., the Euclidean distance), while multi-view subspace clustering often studies subspace representations with the self-expression of samples.

There have been two existing MVC surveys [2, 3]. The differences between theirs and ours are described as follows. On the one hand, the previous two works mostly focus on reviewing the existing shallow models-based MVC, without the organization about the recent deep representation learning-based ones. In our survey, both shallow representation learning-based MVC and deep representation learning-based MVC can be comprehensively considered from the perspective of representation learning, whose detailed introductions can be seen in the following parts. On the other hand, the comprehensive research materials of representation learning-based MVC methods are provided in this survey. Particularly, the commonly used multi-view datasets for experimental evaluation are introduced, and the corresponding download link is provided for beginners. In addition, the open source code library about several representative representation learning-based MVC methods as well as non-representation learning-based ones is reorganized for further study and comparison. Except for the review about widespread research on MVC, some open problems, which may be solved in a manner of representation learning, will be discussed for further advancement of MVC. For clarity, the differences between them are shown in Table 1. We expect that after reading, the researchers can not only have a comprehensive view of the development of MVC, but also be inspired in the future work.

The rest of this paper is organized as follows. In Sect. 2, the main notations as well as basic definition of representation learning-based MVC are provided. Additionally, non-representation learning-based MVC algorithms are briefly introduced in Sect. 3. Several categories of shallow representation learning-based MVC methods are introduced in Sect. 4 in which the corresponding general frameworks are provided. In Sect. 5, we review the existing deep representation learning-based MVC methods. Comprehensive research materials are organized in Sect. 6. In Sect. 7, the research direction and further discussion about the development of MVC are introduced. Finally, the paper is concluded in Sect. 8.

2 Preliminaries

Notations. The main notations in this paper are briefly introduced. Specifically, upper case letters (e.g., $\mathbf{B}$) and bold lower case letters (e.g., $\mathbf{b}$) are used for matrices and vectors, respectively. The Frobenius norm of a matrix can be denoted
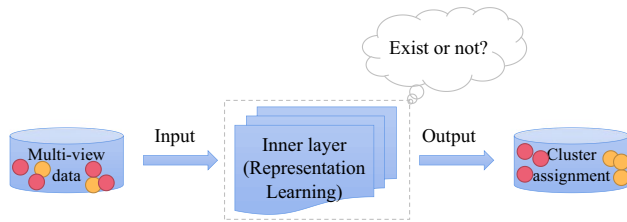
Fig. 2 Illustration of representation learning in MVC

Fig. 3 General framework of co-training style MVC

as $\|\mathbf{B}\|_F = \sqrt{\sum_{i,j} b_{i,j}^2}$, in which $b_{i,j}$ is the entry of $\mathbf{B}$. $\left\{ \mathbf{B}^{(v)} \right\}_{v=1}^{m}$ denotes the corresponding v-th view representation, where m is the number of views. We use $\mathbf{1}$ to denote a column vector with its entries being all 1.

2.1 Description of Representation Learning in MVC

In this paper, representation learning in MVC is defined as such a learning procedure consisting of certain inner layer with the corresponding representation learning in MVC methods, whose general structure is illustrated in Fig. 2. Inspired by [4], it can also be known as the embedding learning or metric learning in MVC methods, especially for multi-view subspace clustering, multi-view spectral clustering, *etc.* In addition, non-representation learning-based ones accordingly make partitions directly from input to output without an extra inner layer, such as the co-training style algorithms, multi-kernel learning, and multi-view *k*-means clustering as well as its variants (multi-view matrix factorization). In the following, before the detailed review of representation learning-based MVC algorithms, non-representation learning-based ones would be briefly introduced.

3 Non-representation Learning-Based Approaches

In this category, we reorganize the literature review into three parts, i.e., co-training style MVC, multi-kernel-based MVC and matrix factorization-based MVC.

Co-training style MVC. Co-training style methods are developed to maximize the mutual agreement across multiple views by training alternately with prior information or knowledge from each other, targeting at achieving the broadest cross-view consensus. The general framework of co-training style MVC is illustrated in Fig. 3. In [5], Bickel et al. first proposed two kinds of MVC frameworks for text data with a co-training strategy. In [6], Zhao et al. seek for discriminative subspaces within a co-training scheme, in which labels automatically learned in one view were deployed to learn discriminative subspace in another view. The work in [7] decomposed different data matrices
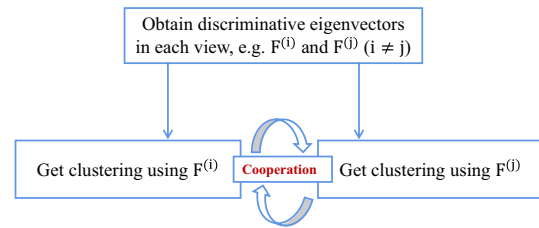
into sparse row and column vectors simultaneously with a minimization algorithm. Liu proposed a guided co-training method for multi-view spectral clustering to deal with large-scale multi-view data, where anchor points were selected to approximate the eigen-decomposition [8]. The work in [9] preserved the global and local structure within the co-training framework. With a co-training schema, Zhao et al. integrated the simplicity of *k*-means and linear discriminant analysis (LDA) [6].

Meanwhile, study has also been focused on the extended versions of co-training [2, 10], i.e., co-EM (Expectation Maximization), co-clustering and co-regularization. The work in [11] developed an automatically weighted multi-view convex mixture method by means of EM. Kumar and Daumé proposed to seek for the clusters which agreed across the multiple views in [12]. The work in [13] co-regularized the clustering hypotheses so that different graphs could agree with each other in the spectral clustering framework. Motivated by the above work, in [14], Ye et al. put forward a co-regularized kernel K-means MVC model, in which the underlying clustering was studied by maximizing the sum of weighted affinities between different clusterings of multiple views. Aiming at clustering objects and features simultaneously, the work in [15] explored a co-clustering model for heterogeneous data, extending the clustering of two modalities from multiple modalities. Bisson and Grimal proposed to learn co-similarities from a collection of matrices, which described the interrelated types of data samples [16]. In [17], Hussain et al. employed the duality in multiple views to deal with multi-view data clustering by co-similarity-based algorithm. The work in [18] combined individual probabilistic latent semantic analysis (PLSA) in multiple views via pairwise co-regularization. Xu et al. simultaneously took two-sided clustering for the co-occurrences along the samples and features into consideration [19]. In [20], Nie et al. developed a co-clustering framework for fast multi-view clustering, in which the data duality between features and samples was considered. In [21], a dynamic auto-weighted multi-view co-clustering was designed to automatically learn weights of different views. Huang et al. constructed a view-specific bipartite graph to extract the co-occurring structure of the data [22].
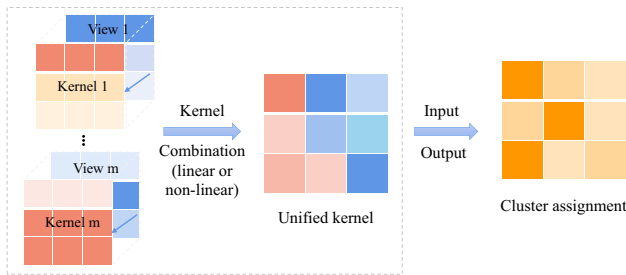
Fig. 4 General framework of multi-kernel-based MVC methods

Multi-kernel-based MVC. Multi-kernel learning methods regard different kernels (e.g., linear kernel, polynomial kernel, as well as gaussian kernel) as multiple views and linearly or non-linearly integrate them to improve the generalization and clustering performance. The general process about multi-kernel learning methods is illustrated in Fig. 4, in which multiple predefined kernels are deployed to represent diverse views. Among these kinds of methods, how to search for suitable kernel functions corresponding to different views and optimally combine them is an important issue.

To deal with the above problem, researchers have made some efforts in this domain. In [23], a custom kernel combination algorithm was designed based on the minimizing-disagreement schema to make full use of information from multiple views. The work in [24] combined multiple kernel with different weights to make full use of all views. Based on kernel alignment, Guo et al. developed a multiple kernel spectral clustering algorithm [25]. In [26], a general multi-view clustering framework was proposed by employing extreme learning machine, where local kernel alignment property was widespread in all views. What is more, a matrix-induced regularization was introduced to reduce the redundancy while enhancing the diversity of different selected kernels [27]. The work in [28] extended the classical k-means algorithm to the kernel space, in which multiple kernel matrices were used to represent data matrices of different views. Likewise, in [29], kernels combination by a localized way designed to learn characteristics of the samples. Houthuys et al. proposed a multi-view kernel spectral clustering method, which was based on a weighted kernel canonical correlation analysis, and suitable to deal with large-scale datasets [30]. The work in [31] explored the multiple kernel clustering via a centered kernel alignment, which could unify the clustering task and multi-kernel learning into a model.

Matrix factorization-based MVC. Matrix factorization is commonly adopted in multi-view clustering, especially non-negative matrix factorization (NMF). NMF decomposes the input non-negative data matrix into two non-negative factor matrices, which is an effective strategy to extract the latent feature by dimensionality-reduction. To some extend,

NMF is inherently correlated with k-means algorithm, whose general formulation can be defined as,

$$\min_{\mathbf{U}^{(v)}, \mathbf{V}^{(v)}} \sum_{v=1}^{m} \|\mathbf{X}^{(v)} - \mathbf{U}^{(v)}\mathbf{V}^{(v)}\|_F^2 + \Psi\left(\mathbf{V}^{(v)}, \mathbf{V}^*\right), \tag{1}$$
$$\text{s.t.} \quad \mathbf{U}^{(v)}, \mathbf{V}^{(v)}, \mathbf{V}^* \geq 0,$$

where $\mathbf{X}^{(v)}$ represents the input multi-view observation. $\mathbf{U}^{(v)}$ is the view-specific centroid matrix, and $\mathbf{V}^{(v)}$ stands for the corresponding cluster assignment matrix. $\Psi(\cdot)$ is certain fusion function to combine multiple $\mathbf{V}^{(v)}$ to obtain the final consensus cluster assignment $\mathbf{V}^*$.

Accordingly, some efforts have been made on this study. Aiming at multi-view data, NMF-based MVC methods attempt to factorize each data matrix into view-specific centroid representation and consensus cluster assignment matrix. In [32], Gao et al. proposed a joint matrix factorization framework, where partition of each view was constrained towards the common clusterings. In [33], the locally geometrical structure of the manifolds was preserved for MVC. A graph regularization was designed and investigated for NMF-based MVC [34]. Zhang et al. proposed a NMF-based model, which enhanced the useful information from multi-view data by means of graph embedding and removed the redundant information [35]. In [36], geometric structures of multiple view data were preserved in both data manifold and feature manifold. Xu et al. embedded multiple discriminative representations into a unified multi-view k-means clustering framework [37]. In [38], Zhang et al. proposed to recover a common cluster structure while preserving the diversity and heterogeneous information of different views in a collaborative manner. Cai et al. proposed a semi-supervised multi-view clustering approach with constrained NMF [39], in which a view-specific representation as well as shared label constraint matrix was constructed. Further, the work in [40] studied a non-negative matrix factorization with orthogonality constraints, focusing on discovering the intra-view diversity as well as a set of independent bases. Liu et al. introduced a weighting scheme in multi-view NMF framework to consider different confidences of multiple views to the consensus results [41]. Recently, in [42], the locally geometrical structure of the data space was retained, and common clustering solution was simultaneously learned while taking different weights of multiple views into account. Since a higher view weight could not ensure the clusters in this view have higher confidences than them in other views, Liu et al. developed a cluster-weighted kernel k-means method for MVC to tackle the issue [43].

Except for NMF-based MVC, there are also some variants, i.e., convex and semi-NMF [44], G-orthogonal NMF [45], and multi-layer NMF [46]. In [47], a regularized semi-NMF is proposed to learn the view-specific low-dimensional graph. For large-scale multi-view data, a

robust multi-view *k*-means clustering method was developed in [48]. Zhao et al. presented a deep matrix factorization framework for MVC based on the semi-NMF strategy, in which the hierarchical semantics of multi-view data were studied via a layer-wise fashion [49]. Considering different contributions about different views toward the consensus graph, a self-weighted schema was designed in the hierarchical framework [50]. In [51], multiple clusterings were discovered by means of a deep matrix factorization-based solution, where each clustering was generated in each layer. Zhang et al. jointly combined the deep matrix decomposition and partition alignment into a unified framework [52].

Other MVC approaches. Except for the co-training style MVC, multi-kernel-based MVC and matrix factorization-based MVC algorithms, there are also some other non-representation learning-based MVC approaches (mostly about the study of multi-view spectral embedding). Based on the bipartite graph, Li et al. proposed a novel multi-view spectral clustering for large-scale datasets, in which local manifold fusion was adopted to integrate heterogeneous information [53]. To explore the block diagonal structure of eigenvectors from the normalized Laplacian matrix, Lu et al. proposed a convex relaxation of sparse spectral clustering method [54]. Yu et al. designed a novel self-paced learning regularizer to assign different weights to multiple views in a multi-view spectral clustering framework [55]. In [56], multi-view spectral clustering was combined with alternative clustering through kernel-based dimensionality reduction to tackle the problem of poor agreement between different views. In [57], Nie et al. made a multi-view extension of the spectral rotation technique, in which view weights were adaptively assigned according to their diverse clustering capacity. With the strategy of brainstorming process, Son et al. developed a spectral clustering method for multi-view data, in which an object was regarded as agenda to be discussed, while we treated each view as a brainstormer [58]. Li et al. proposed to consider the diversity and consistency in both the clustering label matrix and data observation [59]. Chen et al. developed a constrained multi-view spectral clustering approach, where pairwise constraints were explicitly imposed on the unified indicator [60]. To utilize the spatial structure information and the complementary information between different views, the work in [61] characterized the consistency among multiple indicator matrices by means of the weighted tensor nuclear norm. Hu et al. jointly studied the consistent non-negative embedding as well as multiple spectral embeddings [62]. In [63], the spectral embedding and spectral rotation were jointly studied. Hajjar et al. integrated a consensus smoothness and the orthogonality constraint over columns of the non-negative embedding [64].
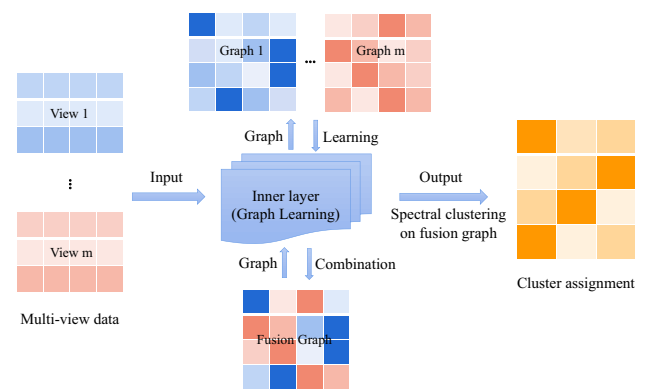

Fig. 5 General framework of multi-view graph clustering methods

4 Shallow Representation Learning-Based MVC

Compared with the non-representation learning-based MVC methods, representation learning-based ones are composed of an inner layer about representation recovery, based on which the corresponding cluster assignment can be subsequently obtained. Up to now, most of the existing studies about MVC are representation learning-based methods. In this part, the shallow and conventional representation learning-based MVC methods are focused, which can be divided into two main categories, i.e., multi-view graph clustering and multi-view subspace clustering. In the following, more detailed information about them will be, respectively, provided.

4.1 Multi-view Graph Clustering

In general, graphs are widely employed to represent the relationships between different samples, in which each node stands for a data sample, and each edge shows the relationship or affinity between pairwise samples. Since there are multiple view observations, multiple graphs are adopted to describe the relationships between different data in multiple views. As mentioned before, only partial information can be captured from the individual view, and accordingly, the individual graph learned separately from each view is not sufficient to comprehensively depict the intact structure of multi-view data. Meanwhile, with the consensus principle, multiple graphs share a common underlying cluster partition of the data. Therefore, to better discover the underlying clustering structure, graphs are eager to be efficiently integrated to mutually reinforce each other. For clarify, the general process about multi-view graph clustering methods is illustrated in Fig. 5, in which a fusion graph is learned across multiple view-specific graphs, and to obtain the final clustering assignment matrix, graph-cut methods or other algorithms (e.g., spectral clustering) are utilized. In particular, we can

define the general formulation of multi-view graph clustering as

$$\min_{\mathbf{A}^{(v)},\mathbf{S}} \sum_{v=1}^{m} \sum_{i \neq j}^{n} \mathrm{Dis}(\mathbf{x}_i^{(v)}, \mathbf{x}_j^{(v)})\mathbf{A}^{(v)} + \lambda\Psi\big(\mathbf{S}, \mathbf{A}^{(v)}\big),$$ (2)

$$\text{s.t.} \quad \mathbf{A}^{(v)} \geqslant 0, \mathbf{a}_i^{(v)}\mathbf{1} = 1, \mathbf{S} \geqslant 0, \mathbf{s}_i\mathbf{1} = 1,$$

where $\mathbf{A}^{(v)}$ is the view-specific affinity graph, and $\mathbf{S}$ represents the consensus similarity matrix by certain fusion of multiple affinity graphs. $\mathrm{Dis}(\cdot)$ stands for some similarity or distance metrics (e.g., the Euclidean distance).

Two types of graph learning are adopted, i.e., affinity/similarity graph, and bipartite graph. For the study of bipartite graph learning, one of the earlier studies proposed by [65] created a bipartite graph with the "minimizing-disagreement" strategy. Based on anchor graphs, Qiang et al. proposed a fast multi-view discrete clustering algorithm, where representative anchors, anchor graphs of multiple views and discrete cluster assignment matrix were constructed with a small time cost [66]. Li et al. learned a consensus bipartite graph for MVC, where the base partitions from multiple views were learned firstly, and then, the common bipartite graph for final clustering was constructed by fusing these base partitions [67]. For the study of similarity graph learning, Nie et al. studied a Laplacian rank constrained graph, which was approximated with each fixed graphs with different confidences [68]. Similarly, with input graphs, Tao et al. developed a MVC framework by a common adaptively learned graph, where the contributions of multiple graphs from view level as well as the performance of affinities in a view from sample-pair level were jointly considered [69]. Liu et al. jointly took both the intra-view relationship and the inter-view correlation into account [70].

Based on the graph-regularized concept factorization, the work in [71] proposed a multi-view document clustering algorithm, where the local geometrical structure of manifolds was preserved. Wang et al. decomposed each low-rank representation into the underlying clustered orthogonal representation for multi-view spectral clustering, in which the orthogonal clustered representations as well as local graph structures shared the same magnitude [72]. In [73], Zhan et al. proposed to learn the optimized (ideal) graphs of multiple views and obtain a global graph by integrating these optimized graphs. The work in [74] explicitly formulated the multi-view consistency and multi-view inconsistency into a unified framework, where each individual graph consisting of the consistent and inconsistent parts and the unified graph were learned. In [75], a general graph-based system for MVC was built, in which data graphs were constructed with different automatically learned weights. Liu et al. attempted to assign nodes with weights and learn the corresponding class label via the nearest neighbor classifier-like

fashion [76]. With a priori rank information, a multi-view Markov chain spectral clustering was developed, in which both the local and global graph embedded in multi-view data were deployed, and the priori knowledge on the rank of the transition probability matrix was incorporated [77]. An effective and parameter-free method to learn a unified graph for MVC by means of cross-view graph diffusion was developed, in which view-specific improved graph was recovered by an iterative cross-diffusion procedure[78]. In [79], Wang et al. proposed a general graph-based MVC method, where data graph matrices of all views were studied, and a unified graph with rank constraint was generated by fusing them. The work in [80] aimed to recover a consensus affinity graph with sparse structures from multiple views.

With the help of emerging tensor technique for the discovery of higher-order correlations among views, some tensor-based MVC methods have been proposed. Given graphs constructed from different views, the work in [81] attempted to integrate higher-order relationships of data by building cross-view tensor product graphs. Ma et al. proposed a MVC framework for graph instances with graph embedding, in which the multi-view graph data were modeled as tensors, and tensor factorization was applied to learn multiple graph embeddings [82]. The work in [83] constructed an essential tensor according to multiple transition probability matrices of the Markov chain. Recently, Wu et al. proposed a unified graph and low-rank tensor learning for MVC, in which each view-specific affinity matrix was learned according to the projected graph learning, and to capture the higher-order correlations among views, the learned affinity graphs were reorganized into the tensor form [84]. The work in [85] explored the view-similar between graphs of multiple views via the minimization of the tensor Schatten p-norm, and view-consensus graph via adaptively weighted strategy. Similarly, to well integrate complementary information and deal with the scalability issue, Jiang et al. exploited a small number of anchor points to construct a tractable graph for each view and then explored the complementary information embedded in these graphs via the tensor Schatten p-norm minimization [86]. A robust affinity graph learning for MVC was proposed in [87], where the robust graph Laplacian from each view was constructed, and then, they were fused in a way to better match the clustering tasks. The work in [88] unified multiple low-rank probability affinity matrices stacked in a low-rank constrained tensor and the consensus indicator graph reflecting the final performance into a framework. To well exploit the complementary information and higher-order correlations among views, a multi-view spectral clustering framework with adaptive graph learning and tensor schatten p-norm minimization was developed, where both the local and global structures were focused [89].

To avoid the curse of dimensionality, the work in [90] developed two parameter-free weighted multi-view projected

clustering frameworks, in which structured graph learning and dimensionality reduction were simultaneously performed, and the obtained similarity matrix directly depicted the clustering indicators. In [91], Wong et al. attempted to describe the noise of each graph with specific structure to clean multiple noisy input graphs. Gao et al. jointly combined the dimensionality reduction, manifold structure learning and feature selection into a unified model, in which the projection matrix was learned to reduce the affection of redundancy features, and the optimal graph with a rank constraint was obtained [92].

Additionally, Rong et al. proposed a multi-metric similarity graph refinement and fusion method for MVC, in which multiple similarity graphs were constructed by different metrics, and a informative unified graph would be learned by adaptively fusing these similarity graphs [93]. Xiao et al. put forward a MVC framework based on knowledge graph embedding, where semantic representations of knowledge elements were generated [94]. Since the number of views increases at each time in the practical applications, Yin et al. proposed an incremental multi-view spectral clustering approach via the sparse and connected graph learning, where a consensus similarity matrix was used to store the structural information of the historical views, and it would be reconstructed when comes the newly collected view [95].

Aiming at the multi-view attributed graph data, a generic framework was designed for clustering them, where a consensus graph regularized by graph contrastive learning was learned [96]. To jointly exploit the attribute and graph structure information, Lin et al. applied a graph filter on features to acquire a smooth representation, selected a few anchor points to construct the corresponding anchor graph, and designed a new regularization term to explore the high-order neighborhood information [97]. The work in [98] exploited the multi-view binary learning for clustering, where the graph structure of data and complementary information of multiple views was combined by the binary code learning. In [99], Zhong et al. integrated consensus graph learning and discretization into a unified framework to avoid the subsequent sub-optimal clustering performance, in which the similarity graph as well as the discrete cluster label matrix was studied.

4.2 Multi-view Subspace Clustering

Assuming that multiple input views are generated from a latent subspace, subspace learning methods target at recovering the latent subspace from multiple views. Multi-view subspace clustering methods, a hot direction in the field of MVC, usually either learn a shared and unified subspace representation from multiple view-specific subspaces or discover a latent space for high-dimensional multi-view data to reduce the dimensionality, based on which latter subspace
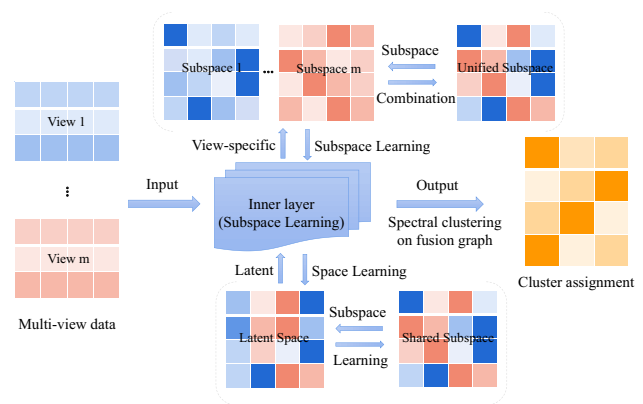

Fig. 6 General framework of multi-view graph clustering methods

learning is conducted. For clarify, the general process about multi-view subspace clustering methods is illustrated in Fig. 6, in which there are commonly two ways of subspace learning. Meanwhile, the objective formulation of multi-view subspace clustering can be commonly rewritten

$$\min_{\mathbf{Z}^{(v)}} \sum_{v=1}^{m} \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)}\|_F^2 + \lambda\Omega(\mathbf{Z}^{(v)}) + \gamma\Psi(\mathbf{Z}, \mathbf{Z}^{(v)}),$$

$$\text{s.t.} \quad \mathbf{Z}^{(v)} \geqslant 0, \mathbf{Z}^{(v)}\mathbf{1} = 1,$$

(3)

where $\mathbf{Z}^{(v)}$ denotes the view-specific subspace representation, and $\mathbf{Z}$ is the learned common representation. $\Omega(\cdot)$ stands for the certain regularization term about $\mathbf{Z}^{(v)}$. To discover the latent space, there is another objective function which can be formulated as

$$\min_{\mathbf{Z}} \sum_{v=1}^{m} \mathcal{F}(\mathbf{X}^{(v)}, \mathbf{H}) + \lambda\|\mathbf{H} - \mathbf{HZ}\|_F^2,$$

$$\text{s.t.} \quad \mathbf{Z} \geqslant 0, \mathbf{Z1} = 1,$$

(4)

where $\mathbf{H}$ denotes the latent space learned from multiple views. After that, subspace learning is conducted on the consensus $\mathbf{H}$. In the following, the existing works would be mainly reviewed via these two means.

On the one hand, to study multiple view-specific subspaces and explore the relationships between them, Li et al. considered each subspace as a state in a Markov chain, which were then sampled via the Monte Carlo methods [100]. To guarantee the consistency among different views, clustering was performed on the subspace representation of each view simultaneously instead of learning the common subspace representation of multiple views [101]. For face images, Zhang et al. proposed to search for an unified subspace structure as a global affinity matrix via integration of multiple affinity matrices in multi-view subspace clustering [102]. The work in [103] exploited the Hilbert Schmidt Independence Criterion (HSIC) as a diversity term to recover the

relationships of multiple subspace representations. In [104], the global low-rank constraint as well as the local cross-topology preserving constraints was considered, by which the learned subspace representations were constrained to be low-rank, and local structure information could be captured. A novel angular-based regularization term was designed for the consensus of multiple views [105]. To locally preserve the data manifold for multi-view subspace clustering, a graph Laplacian was constructed to maintain the view-specific intrinsic geometrical structure, and a sparsity constraint was imposed on the common subspace to better represent the correlations between different data points [106]. Complementary information between views and consensus indicator was jointly considered in [107], where a position-aware exclusivity term was included to deploy the complementary information, and a consistency term was imposed to obtain the consensus indicator. To simultaneously explore the consistency as well as the diversity of multiple views, Luo et al. formulated the construction of multi-view self-representations by a common representation and a set of specific representations [108]. A split multiplicative multi-view subspace clustering algorithm was proposed in [109], in which valuable components from view-specific structures consistent with the consensus subspace representation were extracted. Kang et al. developed a large-scale multi-view subspace clustering algorithm, where smaller graphs from different views were learned [110]. To handle high-dimensional data, the work in [111] introduced a special selection matrix and learned the view weights via their compactness of the cluster structure. To discover the non-linear correlations between multi-view data, some kernel-based multi-view subspace clustering methods are developed. In [112], Zhang et al. proposed a robust low-rank kernel multi-view subspace clustering method, where the non-convex Schatten p-norm regularization term was combined with the kernel trick. The work in [113] designed a one-step kernel multi-view subspace clustering, in which an optimal common affinity matrix was learned.

On the other hand, to discover the shared latent representation, in [114], Zhang et al. sought the latent representation, based on which the data subspace was reconstructed. Further, a generalized version was proposed via neural networks to deal with more general relationships [115]. Tao et al. recovered the row space of the latent representation by assuming that multi-view observations were generated from an latent representation [116]. The work in [117] constructed the global graph and the clustering indicator matrix based on the learned potential embedding space in a unified framework. To be more effective and efficient, Chen et al. further relaxed the constraint of the global similarity matrix, considering more correlations between pairwise data samples [118]. A flexible multi-view representation was learned by enforcing it to be close to multiple views, where the HSIC was effectively exploited [119]. The joint feature selection as well as self-representation learning was combined in a novel MVC framework [120], in which the latent cluster structure was denoted as a block diagonal self-representation. Based on the intact space learning technique, Wang et al. tried to construct an informative intactness-aware similarity representation, which was further constrained by the HSIC to maximize its dependence with the latent space [121]. In [122], the block diagonal constraints were imposed on the fused multi-view features to improve the representation ability. Based on the product Grassmann manifold, a low-rank model for high-dimensional multi-view data clustering was designed, where the low-rank self-representation was learned from product manifolds [123]. Sun et al. combined the anchor learning as well as graph construction into a framework, avoiding the separation of heuristic anchors sampling and clustering process [124].

Similar to the study of multi-view graph clustering, some tensor-based methods are designed to discover the higher-order correlations among views. A low-rank tensor constrained multi-view subspace clustering was first proposed in [125], where each learned subspace representation could be obtained by unfolding the tensor along the corresponding mode. Feng et al. extended the multi-view subspace learning algorithm to the kernel space, in order to learn a more robust local representation [126]. The work in [127] attempted to ensure the consensus among multiple views by introducing a novel type of low-rank tensor constraint on the rotated tensor. Further, in [128], a weighted tensor-nuclear norm based on tensor-singular value decomposition was designed, by which singular values would be regularized with their confidences. In [129], the tensor Schatten-p norm was employed to explore the global low-rank structure of multiple subspace representations, combined with the hyper-Laplacian term. To deal with the tensor nuclear norm minimization problem, Sun et al. proposed a tensor log-regularizer to better approximate the tensor rank [130]. Chen et al. integrated the graph regularized low-rank representation tensor stacked by view-specific subspace and affinity matrix into a unified framework [131]. A generalized nonconvex low-rank tensor approximation was proposed in for multi-view subspace clustering, where physical meanings of multiple singular values were well considered [132].

4.3 Other MVC Algorithms

In addition, there are some other methods based on feature selection or certain metric learning for MVC. A novel unsupervised feature selection method for MVC was developed in [133], where local learning was employed to learn pseudo-class labels on the raw features. Xu et al. performed multi-view data clustering and feature selection simultaneously for high-dimensional data [134]. The work in [135]
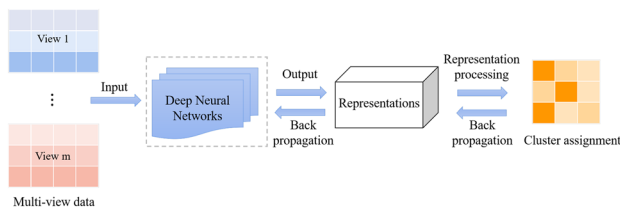
Fig. 7 General framework of multi-view deep representation-based clustering methods

jointly considered the feature learning and partially constrained cluster labels learning, in which feature selection was embedded to select features in each view. The Jaccard similarity was exploited in [136] to measure the clustering consistency across multiple views, where the within-view clustering quality to cluster multi-view data was also considered. A multi-view and multi-exemplar fuzzy clustering method was proposed in [137], where the exemplar of a cluster from one view observation is also an exemplar of that cluster from the other ones. The work in [138] encoded multi-view data into a compact and consistent binary code space and performed binary clustering. Yu et al. designed a fine-grained similarity fusion strategy to acquire the final consensus matrix by considering different view weights [139].

5 Deep Representation Learning-Based MVC

Motivated by the promising progress of deep learning in unsupervised problems, many recent works have been focused on the deep representation learning-based MVC. Specifically, deep representation-based MVC methods implement the deep neural networks (DNNs) as the non-linear parametric mapping functions, as shown in Fig. 7, from which the non-linearity properties that embedded in the original data spaces are fully explored.

Autoencoder is one of the most widespread unsupervised representation learning algorithm. It is composed of encoder and decoder, where the former one maps the data into the latent space with latent representations and the latter reconstructs input data from latent representations. Inspired by [140], Abavisani et al. designed a deep multi-view subspace clustering method based on the autoencoder [141]. Bai et al. investigated and promoted the complementary semantic information of the high-dimensional samples with the enhanced semantic mappings [142]. Xu et al. employed collaborative training scheme with multiple autoencoder networks to better mine the complementary and consistent information of each view, which enabled the latent representations of each view to contain more comprehensively

information. [143]. Moreover, Yang et al. extended the above method with heterogeneous graph learning, which fused the latent representations of different views with adaptive weights [144]. Wang et al. characterized the local and global structure of samples with the representation discriminative constraint, which enforced the common representations of the same cluster to achieve more compact distribution [145]. Zheng et al. utilized the 1-*st* order and the 2-*nd* order graph information to learn a robust local graph of each view and then fused them into self-expressive layers for a consistent subspace representation [146]. In [147], Lu et al. considered both consensus and view-specific representations of each view. All of these representations were concatenated with attention mechanism [148] into a consensus representation matrix for self-representation learning. HSIC [149] was employed by Zhu et al. [150], which ensured the diversities between representations generated by diversity autoencoder net of each view. Meanwhile, samples from all views were regarded as input of one universality autoencoder net for view common representation. Aiming to explore the clustering-friendly features, Cui et al. [151] employed the spectral clustering to generate pseudo-labels as the self-guidance of multi-view encoder fusion layer. Further, Ke et al. leveraged classification loss that calculated by classifier and *k*-means with consensus representation as the self-guidance, which enabled the attention fusion layer to adaptively assign high weights to clustering-friendly representations generated by encoders of each view in a back-propagation manner [152]. Zhang et al. integrated the pseudo-label learning, view consensus representation learning and view-specific representation learning into an end-to-end framework [153]. Du et al. aligned representations from different views with the discrepancy constraint [154]. To solve the issue that representation alignment might result in less separable clusters of representation, in [155], Trosten et al. introduced the view-wise contrastive learning. Wang et al. explored a multi-view fuzzy clustering method that adopted the joint learning of deep random walk and sparse low-rank embedding [156].

To fully leverage the features embedded in the attributed multi-view graph data, graph neural network (GNN) [157] was applied to deep representation-based MVC. Fan et al. chose a most reliable view as the input of one graph convolutional network (GCN) [158] encoder and multiple GCN decoder to capture the view-consistent low-dimensional feature representation among different views [159]. In [160], the dual encoders were introduced to explore the graph embedding and consistent embedding of high-dimensional samples, in which the GCN encoders were designed to explore the view-specific graph embedding and a consensus clustering embedding among all views could be further obtained via consistent embedding encoders. In [161], Zhang et al. leveraged multiple GCN-based autoencoders to mine the local structures and fused them into a global

representation which was further processed by a variational graph autoencoder [162] with nearest neighbor constraint for deep global latent representation. Cai et al. leveraged the global GCN autoencoder and the partial GCN autoencoders to mine the global structure and unique structures of different views simultaneously and then fused them into a self-training clustering module with adaptively weighted fusion [163]. In [164], Xia et al. imposed the diagonal constraint on the consensus representation that generated by multiple GCN autoencoders with the self-expression scheme for better clustering capability.

Additionally, impressive clustering performance can also be achieved by the methods based on generative adversarial networks (GAN) [165]. In [166], GAN were utilized to reconstruct the samples from a common representation that shared by multiple views. Further, they took the differences between views into consideration and extended the above method with a fusion module, which fused multiple representations from different views with adaptive weights [167]. In [168], Zhou et al. aligned the latent representations of multi-encoders with the adversarial learning and fused them into a clustering layer with an attention mechanism to obtain more compact cluster partitions. In [169], Li et al. partitioned representations that generated by encoders into view-specific features and view-consistent features, where a adversarial learning process was developed to enable view-consistent features to be non-trivial. In [170], Huang et al. incorporated the deep neural networks with multi-view spectral clustering, where the local invariance of each view and the view-consistent representation among different views were optimized in an unified framework.

Apart from the autoencoder as well as GNN-based deep representation learning for MVC methods, some other attempts have also been made by researchers. In [171], a contrastive fusion module was employed, enabling the representation fusion layer to generate more clustering relevant representations. Xin et al. maximized canonical correlation between views for robust consensus representations of each view [172]. Xu et al. employed variational autoencoders (VAEs) [173] to disentangle the representations of each view into view-specific visual representations and view-consensus cluster representations [174]. Joint learning for deep MVC was introduced in [175], where the representations extracted by the stacked autoencoder, convolutional autoencoder as well as convolutional VAE were fused by multi-view fusion modules to explore the clustering-friendly representations.

6 Research Materials

In this section, some widely used multi-view datasets as well as open source codes of several representative MVC methods are summarized. To be specific, according to different entries multi-view datasets described, we divide the multi-view datasets into three categories, i.e., the image, document and graph datasets. The statistics of them are, respectively, introduced as follows. For some representative MVC methods reviewed before, the source code links are provided as well, and all summarized multi-view datasets can be directly downloaded at https://github.com/ManshengChen/Multi-view_Datasets.

6.1 Image Multi-view Datasets

Most of the image multi-view datasets are generated based on the original single-view image datasets. We would introduce some popular pre-computed multi-view image datasets. Readers can also generate the similar multi-view datasets by yourself for further research.

Yale[1] is a widely used gray face dataset, which contains 15 categories and each category has 11 images. Three image features, i.e., the intensity feature (4096-D), the LBP feature [176] (3304-D) and the Gabor feature [177] (6750-D) are widely used as three types of views. More details are provided in the previous work [125].

ORL consists of 40 distinct objects and provides 10 different images for each object, which is captured under different times, lighting, facial expressions, and facial details. The three types of view are same as Yale.

Caltech101[2] contains 8677 images of 101 categories. Each category has 40 to 100 images. There are three popular multi-view datasets generated from Caltech101, which are, respectively, Caltech101-7, Caltech101-20 and Caltech101. The Caltech101-7 and Caltech101-20 datasets with the size of 1474 and 2386 are generated from the subset of Caltech101, where 7 and 20 categories are, respectively, included. Both of them have six views consisting of the Gabor feature (48-D), the Wavelet-moment feature [178] (40-D), the Centrist feature [179] (254-D), the HOG feature [180] (1984-D), the GIST feature [181] (512-D), and the LBP feature (928-D). The Caltech101 [127] dataset is composed of four views, whose dimensions are, respectively, 2048, 4800, 3540,1240.

Notting-Hill is extracted from the movie "Notting-Hill" and contains 4660 facial images belonging to 5 classes. Each image is represented by three kinds of features, i.e., the Gabor feature (6750-D), the LBP feature (3304-D), and the intensity feature (2000-D).

Columbia Object Image Library (COIL20)[3] contains 1440 images of 20 categories. Three kinds of features were used to describe the images, which are the intensity feature

---

[1] http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[2] http://www.vision.caltech.edu/Image_Datasets/Caltech101/.

[3] http://www.cs.columbia.edu/CAVE/software/softlib/.

(1024-D), the LBP feature (3304-D), and the Gabor feature (3304-D).

Multiple feature handwriting digit dataset (Mfeat) records the handwriting digital numbers from 0 to 9, and each number class has 200 handwriting digital images. The profile correlation feature (216-D), Fourier coefficient feature [182] (76-D), Karhunen–Loeve coefficient feature [183] (64-D), morphological feature (6-D), pixel average feature (240-D), and Zernike moment feature [184] (47-D) are selected to represent an image.

Hdigit is a handwriting digit dataset from MNIST Handwritten Digits and USPS Handwritten Digits. This dataset consists of 10,000 samples from two views, whose dimensions are 784 and 256, respectively.

MSRC-v1 is an image dataset containing seven classes. Each image is represented by four different kinds of features, i.e., the Color moment feature (24-D), the HOG feature (576-D), the LBP feature(256-D) and the CENT feature (254-D).

Scene-15 is a scene image dataset with 15 categories. It has 4485 images and represents each image with the Pyramid Histograms Of visual Words (PHOW) feature (1800-D), the Pair-wise Rotation Invariant Co-occurrence Local Binary Pattern (PRI-CoLBP) feature (1180-D) and the Centrist feature (1240-D).

One-hundred plant species leaves data set (100leaves) has 1600 samples from 100 plant species. For each species, the shape descriptor, fine scale margin and texture histogram are selected as three views of the features.

6.2 Document Multi-view Datasets

The document multi-view datasets are generated from single view dataset or naturally described in multiple views. For instance, news can be reported by several organizations, which could be regarded as features from several views. Some commonly used document multi-view datasets are introduced as follows.

BBC4view[4] is collected from the BBC news website. The dataset consists of 685 documents belonging to five topical areas. Four kinds of features are used, whose dimensions are 4659, 4633, 4665 and 4684, respectively.

BBCSport[5] contains 544 BBC news about sport in 5 classes. Three kinds of views are selected, whose dimensions are 3183 and 3203, respectively

WebKB is a web page dataset containing 203 web pages in 4 classes. The content of the page, anchor text of the hyper-link, and text in its title are selected as three kinds of feature.

3sources[6] consists of 169 news reported by three news organizations, i.e., BBC, Reuters, and Guardian. These news are classified into six categories according to their topics.

Reuters is a multiple language document dataset. It has 18758 documents with 6 classes. Each document is written in five languages, i.e., English, French, German, Spanish and Italian.

Citeseer contains 3312 scientific papers with six classes. There are two kinds of features, i.e., links and word presence information.

Cora consists of 2708 documents with seven classes (i.e., probabilistic methods, rule learning, reinforcement learning, neural networks, theory, genetic algorithms, and case- based methods). Two kinds of features are often selected, i.e., content and cites.

6.3 Graph Multi-view Datasets

The graph multi-view datasets contain one view of attribute features about samples and multiple views of nearest neighbor graphs describing the relationships between them.

ACM is a paper graph dataset with 3,025 papers. The attribute features are extracted from paper keywords, and two graphs are constructed based on the co-author and co-subject relationships. Co-author represents two papers are written by the same author and co-subject suggests that they focus on the same research area.

DBLP is an author graph dataset with 4,057 authors. The attribute features are constructed based on the authors' keywords. Three graphs are generated from three relations, i.e., co-author, co-conference and co-term. Co-author indicates two authors work together on the same papers, co-conference represents two authors publish papers in the same conference, and co-term means that two authors publish papers with the same term.

IMDB is a movie graph dataset from the IMDB with size of 4708. Two graphs are constructed based on the co-actor and co-director relationships. Co-actor means two movies are acted by the same actor, and co-director represents two movies are directed by the same director.

6.4 Open Source Code Library

The open source codes of several MVC methods are summarized, including two non-representation learning-based MVC methods and fourteen representation learning-based MVC methods. The corresponding code links are presented in Table 2.

Table 2 Collection of open source code

Category	Method	Code Link
Non-rep.	Co-train [12]	www.abhishek.umiacs.io/code_cospectral.zip
	Co-reg [13]	www.abhishek.umiacs.io/code_coregspectral.zip
Rep.	GMC [79]	https://github.com/cshaowang/gmc
	MVGL [73]	https://github.com/kunzhan/MVGL
	MvAGC [97]	https://github.com/sckangz/MvAGC
	t-SVD-MSC [127]	www.researchgate.net/profile/Yuan_Xie4
	ETLMSC [83]	https://github.com/Jlwu1992/Code-for-ETLMSC
	LTBPL [88]	https://github.com/ManshengChen/Code-for-LTBPL
	LMVSC [110]	https://github.com/sckangz/LMVSC
	MCLES [117]	https://github.com/ManshengChen/MCLES
	DEMVC [143]	https://github.com/SubmissionsIn/DEMVC
	MSCNLG [146]	https://github.com/ZQH92/MSCNLG
	MvDSCN [150]	https://github.com/huybery/MvDSCN
	O2MAC [159]	https://github.com/googlebaba/WWW2020-O2MAC
	CONAN [171]	https://github.com/Guanzhou-Ke/conan
	Multi-VAE [174]	https://github.com/SubmissionsIn/Multi-VAE

7 Research Direction and Discussion

Deep learning-based MVC. Deep learning technique has achieved impressive performance in many research areas, such as image segmentation and object detection. Compared with the traditional shallow models, deep learning-based ones exhibit better expression ability, which may help to recover more underlying semantic information. Unfortunately, most of them are limited in the study of supervised learning. For unsupervised deep representation learning, there is still a lot of room for development, particularly deep multi-view clustering. More detailed information can be reviewed in 4. Although some existing studies have been proposed to learn the embedding representation, based on which clustering is conducted, from deep multi-view representation learning, there should be more advanced strategies to well explore the correlations between deep representation learning and multi-view clustering. In the meantime, theoretical investigation should also be focused to show why the deep representation learning-based MVC methods can achieve such excellent performance.

Large Scale MVC. In the big data era, there are a quantity of practical multi-view data with large size or dimension. However, most of the existing works are merely available for smaller datasets, due to the computation of an $n \times n$ graph representation as well as eigen-decomposition. Note that in the real-world applications, it is essential for researchers to design efficient algorithms to deal with the large-scale data. Recently, some attempts have been made to handle this challenging problem, most of which are based on the anchor representation learning. More meaningful study can be devoted into this direction to solve more bigger data (even of million or billion) and improve the clustering performance shared by multi-view data. For large-scale multi-view data clustering with higher dimension, how to effectively and efficiently deal with them still remains a tough problem.

Incomplete MVC. Assuming that all views are complete, MVC methods attempt to make data partitions by fusing the useful information from multiple views. However, in real-world applications, the prior assumption is hard to be achieved. In other words, there would be some randomly missing information from views, leading to the study of incomplete multi-view data clustering. A direct way to deal with this problem is to replace the missing values with zero or mean values. But unfortunately, they fail to take the correlations between different non-missing data points into account, producing not ideal results. In the past few years, several efforts have been made to improve the clustering performance. For instance, Wen et al. proposed a unified embedding alignment framework, in which a locality-preserved reconstruction term was used for inferring missing views, and a consistent graph was learned to guarantee the local structure of multi-view data [185]. Further, they exploited the tensor low-rank representation constraint to recover more reasonable missing views by learning deeper inter-view and intra-view information [186]. Despite of the impressive performance, researchers in this area always conduct experiments on the incomplete datasets randomly generated by themselves, which would exist much uncertainty. Besides, more focus should be put to deal with more bigger incomplete multi-view data according to the emerging anchor representation learning strategy.

Uncoupled MVC. Both of the existing MVC and incomplete MVC methods require that a data sample in a view is completely or partially mapped onto one or more data points in another view, yet in the practical applications, there are

still some unknown mapping relationships between different multi-view data, which leads to the study of uncoupled MVC. Unknown mapped multi-view data represent that no mapping relationships between any pairs of samples are known. Specifically, there are two kinds of conditions, i.e., multi-view data with unknown single mapping relationships as well as multi-view data with unknown complex mapping relationships. Up to now, the work in [187] is the first and unique attempt to deal with this problem of the first condition. More attempts are expected to be made by exploring the cross-view consistency of diverse representations for further study, and more practical unknown mapped multi-view datasets should be collected for experiments.

MVC with mixed data types. In multi-view data, there are not necessarily just numerical or categorical features. Other types (e.g., symbolic or ordinal) can also be contained, which may exist in the same view or different views. A naive strategy to handle them is to convert all of them to the categorical type, which would certainly lose much useful information during the learning procedure. Few attempts have been made to integrate different types of data to conduct MVC. For instance, in [188], vine copulas was exploited to deal with the problem of mixed data types. Wang et al. proposed a convex formalization by integrating all available data sources and made feature selection to avoid the curse of dimensionality [189]. More explorations are eager to be made into the study of MVC with mixed data types to obtain clustering-friendly representations.

8 Conclusion

In this paper, a novel taxonomy is proposed to sort out the existing MVC algorithms, which are mainly divided into two categories, i.e., non-representation learning-based MVC and representation learning-based MVC. Particularly, the representation learning-based MVC methods therein are our focus, which consists of two kinds of learning models to integrate useful information from different views, i.e., shallow representation learning-based MVC and deep representation learning-based MVC. According to different ways of representation learning, shallow models can be further divided into two main groups, i.e., multi-view graph clustering and multi-view subspace clustering. The developing deep models with better expression for more complex data structures are also sorted out in this survey. To have a comprehensive view of MVC, basic research materials of MVC methods are introduced for readers, including introductions of the commonly used multi-view datasets with the download link, and the open source code library of some representative MVC methods. Last but not least, several open and challenging problems are pointed out to inspire researchers for further investigation and advancement.

References

1. Xu C, Tao D, Xu C (2013) A survey on multi-view learning. CoRR arXiv:1304.5634
2. Yang Y, Wang H (2018) Multi-view clustering: a survey. Big Data Min Anal 1(2):83–107
3. Chao G, Sun S, Bi J (2021) A survey on multiview clustering. IEEE Trans Artif Intell 2(2):146–168
4. Bengio Y, Courville AC, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828
5. Bickel S, Scheffer T (2004) Multi-view clustering. In: ICDM, pp. 19–26
6. Zhao X, Evans NWD, Dugelay J (2014) A subspace co-training framework for multi-view clustering. Pattern Recognit Lett 41:73–82
7. Sun J, Lu J, Xu T, Bi J (2015) Multi-view sparse co-clustering via proximal alternating linearized minimization. ICML 37:757–766
8. Liu T (2017) Guided co-training for large-scale multi-view spectral clustering. CoRR arXiv:1707.09866
9. Cai W, Zhou H, Xu L (2021) A multi-view co-training clustering algorithm based on global and local structure preserving. IEEE Access 9:29293–29302
10. Nigam K, Ghani R (2000) Analyzing the effectiveness and applicability of co-training. In: CIKM, pp 86–93
11. Tzortzis G, Likas CL (2010) Multiple view clustering using a weighted combination of exemplar-based mixture models. IEEE Trans Neural Networks Learn Syst 21(12):1925–1938
12. Kumar A, III HD (2011) A co-training approach for multi-view spectral clustering. In: Getoor L, Scheffer T (eds) ICML, pp 393–400
13. Kumar A, Rai P, III HD (2011) Co-regularized multi-view spectral clustering. In: NIPS, pp 1413–1421
14. Ye Y, Liu X, Yin J, Zhu E (2016) Co-regularized kernel k-means for multi-view clustering. In: ICPR, pp 1583–1588
15. Meng L, Tan A, Xu D (2014) Semi-supervised heterogeneous fusion for multimedia data co-clustering. IEEE Trans Knowl Data Eng 26(9):2293–2306
16. Bisson G, Grimal C (2012) Co-clustering of multi-view datasets: a parallelizable approach. In: ICDM, pp 828–833
17. Hussain SF, Bashir S (2016) Co-clustering of multi-view datasets. Knowl Inf Syst 47(3):545–570
18. Jiang Y, Liu J, Li Z, Li P, Lu H (2012) Co-regularized PLSA for multi-view clustering. ACCV 7725:202–213

19. Xu P, Deng Z, Choi K, Cao L, Wang S (2019) Multi-view information-theoretic co-clustering for co-occurrence data. In: AAAI, pp 379–386
20. Nie F, Shi S, Li X (2020) Auto-weighted multi-view co-clustering via fast matrix factorization. Pattern Recognit 102:107207
21. Hu S, Yan X, Ye Y (2020) Dynamic auto-weighted multi-view co-clustering. Pattern Recognit 99
22. Huang S, Xu Z, Tsang IW, Kang Z (2020) Auto-weighted multi-view co-clustering with bipartite graphs. Inf Sci 512:18–30
23. de Sa VR, Gallagher PW, Lewis JM, Malave VL (2010) Multi-view kernel construction. Mach Learn 79(1–2):47–71
24. Tzortzis G, Likas A (2012) Kernel-based weighted multi-view clustering. In: ICDM, pp 675–684
25. Guo D, Zhang J, Liu X, Cui Y, Zhao C (2014) Multiple kernel learning based multi-view spectral clustering. In: ICPR, pp 3774–3779
26. Wang Q, Dou Y, Liu X, Xia F, Lv Q, Yang K (2018) Local kernel alignment based multi-view clustering using extreme learning machine. Neurocomputing 275:1099–1111
27. Liu X, Dou Y, Yin J, Wang L, Zhu E (2016) Multiple kernel k-means clustering with matrix-induced regularization. In: AAAI, pp 1888–1894
28. Yu S, Tranchevent L, Liu X, Glänzel W, Suykens JAK, Moor BD, Moreau Y (2012) Optimized data fusion for kernel k-means clustering. IEEE Trans Pattern Anal Mach Intell 34(5):1031–1039
29. Gönen M, Margolin AA (2014) Localized data fusion for kernel k-means clustering with application to cancer biology. In: NIPS, pp 1305–1313
30. Houthuys L, Langone R, Suykens JAK (2018) Multi-view kernel spectral clustering. Inf Fusion 44:46–56
31. Lu Y, Wang L, Lu J, Yang J, Shen C (2014) Multiple kernel clustering based on centered kernel alignment. Pattern Recognit 47(11):3656–3664
32. Gao J, Han J, Liu J, Wang C (2013) Multi-view clustering via joint nonnegative matrix factorization. In: SDM, pp 252–260
33. Zhang X, Zhao L, Zong L, Liu X, Yu H (2014) Multi-view clustering via multi-manifold regularized nonnegative matrix factorization. In: ICDM, pp 1103–1108
34. Wang Z, Kong X, Fu H, Li M, Zhang Y (2015) Feature extraction via multi-view non-negative matrix factorization with local graph regularization. In: ICIP, pp 3500–3504
35. Zhang X, Gao H, Li G, Zhao J, Huo J, Yin J, Liu Y, Zheng L (2018) Multi-view clustering based on graph-regularized nonnegative matrix factorization for object recognition. Inf Sci 432:463–478
36. Luo P, Peng J, Guan Z, Fan J (2018) Dual regularized multi-view non-negative matrix factorization for clustering. Neurocomputing 294:1–11
37. Xu J, Han J, Nie F (2016) Discriminatively embedded k-means for multi-view clustering. In: CVPR, pp 5356–5364
38. Zhang G, Wang C, Huang D, Zheng W, Zhou Y (2018) Tw-co-k-means: two-level weighted collaborative k-means for multi-view clustering. Knowl Based Syst 150:127–138
39. Cai H, Liu B, Xiao Y, Lin L (2019) Semi-supervised multi-view clustering based on constrained nonnegative matrix factorization. Knowl Based Syst 182
40. Cai H, Liu B, Xiao Y, Lin L (2020) Semi-supervised multi-view clustering based on orthonormality-constrained nonnegative matrix factorization. Inf Sci 536:171–184
41. Liu SS, Lin L (2021) Integrative clustering of multi-view data by nonnegative matrix factorization. CoRR arXiv:2110.13240
42. Khan GA, Hu J, Li T, Diallo B, Wang H (2022) Multi-view data clustering via non-negative matrix factorization with manifold regularization. Int J Mach Learn Cybern 13(3):677–689
43. Liu J, Cao F, Gao X, Yu L, Liang J (2020) A cluster-weighted kernel k-means method for multi-view clustering. In: AAAI, pp 4860–4867
44. Ding CHQ, Li T, Jordan MI (2010) Convex and semi-nonnegative matrix factorizations. IEEE Trans Pattern Anal Mach Intell 32(1):45–55
45. Ding CHQ, He X, Simon HD (2005) Nonnegative lagrangian relaxation of K-means and spectral clustering. ECML 3720:530–538
46. Cichocki A, Zdunek R (2006) Multilayer nonnegative matrix factorisation. Electron Lett 42:947–948
47. Li G, Han K, Pan Z, Wang S, Song D (2021) Multi-view image clustering via representations fusion method with semi-nonnegative matrix factorization. IEEE Access 9:96233–96243
48. Cai X, Nie F, Huang H (2013) Multi-view k-means clustering on big data. In: IJCAI, pp 2598–2604
49. Zhao H, Ding Z, Fu Y (2017) Multi-view clustering via deep matrix factorization. In: AAAI, pp 2921–2927
50. Cui B, Yu H, Zhang T, Li S (2019) Self-weighted multi-view clustering with deep matrix factorization. ACML 101:567–582
51. Wei S, Wang J, Yu G, Domeniconi C, Zhang X (2020) Multi-view multiple clusterings using deep matrix factorization. In: AAAI, pp 6348–6355
52. Zhang C, Wang S, Liu J, Zhou S, Zhang P, Liu X, Zhu E, Zhang C (2021) Multi-view clustering via deep matrix factorization and partition alignment. In: MM, pp 4156–4164
53. Li Y, Nie F, Huang H, Huang J (2015) Large-scale multi-view spectral clustering via bipartite graph. In: AAAI, pp 2750–2756
54. Lu C, Yan S, Lin Z (2016) Convex sparse spectral clustering: single-view to multi-view. IEEE Trans Image Process 25(6):2833–2843
55. Yu H, Lian Y, Zong L, Tian L (2017) Self-paced learning based multi-view spectral clustering. In: ICTAI, pp 6–10
56. He X, Li L, Roqueiro D, Borgwardt KM (2017) Multi-view spectral clustering on conflicting views. ECML 10535:826–842
57. Nie F, Tian L, Li X (2018) Multiview clustering via adaptively weighted procrustes. In: KDD, pp 2022–2030
58. Son JW, Jeon J, Lee A, Kim S (2017) Spectral clustering with brainstorming process for multi-view data. In: AAAI, pp 2548–2554
59. Li Z, Tang C, Chen J, Wan C, Yan W, Liu X (2019) Diversity and consistency learning guided spectral embedding for multi-view clustering. Neurocomputing 370:128–139
60. Chen C, Qian H, Chen W, Zheng Z, Zhu H (2019) Auto-weighted multi-view constrained spectral clustering. Neurocomputing 366:1–11
61. Xu H, Zhang X, Xia W, Gao Q, Gao X (2020) Low-rank tensor constrained co-regularized multi-view spectral clustering. Neural Netw 132:245–252
62. Hu Z, Nie F, Wang R, Li X (2020) Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding. Inf Fusion 55:251–259
63. Wan Z, Xu H, Gao Q (2021) Multi-view clustering by joint spectral embedding and spectral rotation. Neurocomputing 462:123–131
64. Hajjar SE, Dornaika F, Abdallah F (2022) Multi-view spectral clustering via constrained nonnegative embedding. Inf Fusion 78:209–217
65. De Sa VR (2005) Spectral clustering with two views. In: ICML, pp 20–27
66. Qiang Q, Zhang B, Wang F, Nie F (2021) Fast multi-view discrete clustering with anchor graphs. In: AAAI, pp 9360–9367
67. Li M, Liang W, Liu X (2021) Multi-view clustering with learned bipartite graph. IEEE Access 9:87952–87961
68. Nie F, Li J, Li X (2017) Self-weighted multiview clustering with multiple graphs. In: IJCAI, pp 2564–2570

69. Tao H, Hou C, Zhu J, Yi D (2017) Multi-view clustering with adaptively learned graph. ACML 77:113–128
70. Liu B, Huang L, Wang C, Fan S, Yu PS (2021) Adaptively weighted multiview proximity learning for clustering. IEEE Trans Cybern 51(3):1571–1585
71. Zhan K, Shi J, Wang J, Tian F (2017) Graph-regularized concept factorization for multi-view document clustering. J Vis Commun Image Represent 48:411–418
72. Wang Y, Wu L (2018) Beyond low-rank representations: orthogonal clustering basis reconstruction with optimized graph structure for multi-view spectral clustering. Neural Netw 103:1–8
73. Zhan K, Zhang C, Guan J, Wang J (2018) Graph learning for multiview clustering. IEEE Trans Cybern 48(10):2887–2895
74. Liang Y, Huang D, Wang C (2019) Consistency meets inconsistency: a unified graph learning framework for multi-view clustering. In: ICDM, pp 1204–1209
75. Wang H, Yang Y, Liu B, Fujita H (2019) A study of graph-based system for multi-view clustering. Knowl Based Syst 163:1009–1019
76. Liu S, Ding C, Jiang F, Wang Y, Yin B (2019) Auto-weighted multi-view learning for semi-supervised graph clustering. Neurocomputing 362:19–32
77. Xie D, Gao Q, Wang Q, Xiao S (2019) Multi-view spectral clustering via integrating global and local graphs. IEEE Access 7:31197–31206
78. Tang C, Liu X, Zhu X, Zhu E, Luo Z, Wang L, Gao W (2020) CGD: multi-view clustering via cross-view graph diffusion. In: AAAI, pp 5924–5931
79. Wang H, Yang Y, Liu B (2020) GMC: graph-based multi-view clustering. IEEE Trans Knowl Data Eng 32(6):1116–1129
80. Hu Z, Nie F, Chang W, Hao S, Wang R, Li X (2020) Multi-view spectral clustering via sparse graph learning. Neurocomputing 384:1–10
81. Shu L, Latecki LJ (2015) Integration of single-view graphs with diffusion of tensor product graphs for multi-view spectral clustering. ACML 45:362–377
82. Ma G, He L, Lu C, Shao W, Yu PS, Leow AD, Ragin AB (2017) Multi-view clustering with graph embedding for connectome analysis. In: CIKM, pp 127–136
83. Wu J, Lin Z, Zha H (2019) Essential tensor learning for multi-view spectral clustering. IEEE Trans Image Process 28(12):5910–5922
84. Wu J, Xie X, Nie L, Lin Z, Zha H (2020) Unified graph and low-rank tensor learning for multi-view clustering. In: AAAI, pp 6388–6395
85. Gao Q, Xia W, Gao X, Tao D (2021) Effective and efficient graph learning for multi-view clustering. CoRR arXiv:2108.06734
86. Jiang T, Gao Q (2021) Multiple graph learning for scalable multi-view clustering. CoRR arXiv:2106.15382
87. Jing P, Su Y, Li Z, Nie L (2021) Learning robust affinity graph representation for multi-view clustering. Inf Sci 544:155–167
88. Chen M, Wang C, Lai J (2022) Low-rank tensor based proximity learning for multi-view clustering. IEEE Trans Knowl Data Eng
89. Zhao Y, Yun Y, Zhang X, Li Q, Gao Q (2022) Multi-view spectral clustering with adaptive graph learning and tensor schatten p-norm. Neurocomputing 468:257–264
90. Wang R, Nie F, Wang Z, Hu H, Li X (2020) Parameter-free weighted multi-view projected clustering with structured graph learning. IEEE Trans Knowl Data Eng 32(10):2014–2025
91. Wong WK, Han N, Fang X, Zhan S, Wen J (2020) Clustering structure-induced robust multi-view graph recovery. IEEE Trans Circuits Syst Video Technol 30(10):3584–3597
92. Gao Q, Wan Z, Liang Y, Wang Q, Liu Y, Shao L (2020) Multi-view projected clustering with graph learning. Neural Netw 126:335–346
93. Rong W, Zhuo E, Tao G, Cai H (2021) Effective and adaptive refined multi-metric similarity graph fusion for multi-view clustering. PAKDD 12713:194–206
94. Xiao H, Chen Y, Shi X (2021) Knowledge graph embedding based on multi-view clustering framework. IEEE Trans Knowl Data Eng 33(2):585–596
95. Yin H, Hu W, Zhang Z, Lou J, Miao M (2021) Incremental multi-view spectral clustering with sparse and connected graph learning. Neural Netw 144:260–270
96. Pan E, Kang Z (2021) Multi-view contrastive graph clustering. CoRR arXiv:2110.11842
97. Lin Z, Kang Z (2021) Graph filter-based multi-view attributed graph clustering. In: IJCAI, pp 2723–2729
98. Jiang G, Wang H, Peng J, Chen D, Fu X (2021) Graph-based multi-view binary learning for image clustering. Neurocomputing 427:225–237
99. Zhong G, Shu T, Huang G, Yan X (2022) Multi-view spectral clustering by simultaneous consensus graph learning and discretization. Knowl Based Syst 235:107632
100. Li G, Günnemann S, Zaki MJ (2013) Stochastic subspace search for top-k multi-view clustering. In: KDD, p 3
101. Gao H, Nie F, Li X, Huang H (2015) Multi-view subspace clustering. In: ICCV, pp 4238–4246
102. Zhang X, Phung DQ, Venkatesh S, Pham D, Liu W (2015) Multi-view subspace clustering for face images. In: DICTA, pp 1–7
103. Cao X, Zhang C, Fu H, Liu S, Zhang H (2015) Diversity-induced multi-view subspace clustering. In: CVPR, pp 586–594
104. Fan Y, He R, Hu B (2015) Global and local consistent multi-view subspace clustering. In: ACPR, pp 564–568
105. Wang Y, Lin X, Wu L, Zhang W, Zhang Q, Huang X (2015) Robust subspace clustering for multi-view data by exploiting correlation consensus. IEEE Trans Image Process 24(11):3939–3949
106. Wang L, Li D, He T, Xue Z (2016) Manifold regularized multi-view subspace clustering for image representation. In: ICPR, pp 283–288
107. Wang X, Guo X, Lei Z, Zhang C, Li SZ (2017) Exclusivity-consistency regularized multi-view subspace clustering. In: CVPR, pp 1–9
108. Luo S, Zhang C, Zhang W, Cao X (2018) Consistent and specific multi-view subspace clustering. In: AAAI, pp 3730–3737
109. Yang Z, Xu Q, Zhang W, Cao X, Huang Q (2019) Split multiplicative multi-view subspace clustering. IEEE Trans Image Process 28(10):5147–5160
110. Kang Z, Zhou W, Zhao Z, Shao J, Han M, Xu Z (2020) Large-scale multi-view subspace clustering in linear time. In: AAAI, pp 4412–4419
111. Yan F, Wang X, Zeng Z, Hong C (2020) Adaptive multi-view subspace clustering for high-dimensional data. Pattern Recognit Lett 130:299–305
112. Zhang X, Sun H, Liu Z, Ren Z, Cui Q, Li Y (2019) Robust low-rank kernel multi-view subspace clustering based on the schatten p-norm and correntropy. Inf Sci 477:430–447
113. Zhang G, Zhou Y, He X, Wang C, Huang D (2020) One-step kernel multi-view subspace clustering. Knowl Based Syst 189
114. Zhang C, Hu Q, Fu H, Zhu P, Cao X (2017) Latent multi-view subspace clustering. In: CVPR, pp 4333–4341
115. Zhang C, Fu H, Hu Q, Cao X, Xie Y, Tao D, Xu D (2020) Generalized latent multi-view subspace clustering. IEEE Trans Pattern Anal Mach Intell 42(1):86–99
116. Tao H, Hou C, Qian Y, Zhu J, Yi D (2020) Latent complete row space recovery for multi-view subspace clustering. IEEE Trans Image Process 29:8083–8096
117. Chen M, Huang L, Wang C, Huang D (2020) Multi-view clustering in latent embedding space. In: AAAI, pp 3513–3520

118. Chen M, Huang L, Wang C, Huang D, Lai J (2021) Relaxed multi-view clustering in latent embedding space. Inf Fusion 68:8–21

119. Li R, Zhang C, Hu Q, Zhu P, Wang Z (2019) Flexible multi-view representation learning for subspace clustering. In: IJCAI, pp 2916–2922

120. Yan H, Liu S, Yu PS (2019) From joint feature selection and self-representation learning to robust multi-view subspace clustering. In: ICDM, pp 1414–1419

121. Wang X, Lei Z, Guo X, Zhang C, Shi H, Li SZ (2019) Multi-view subspace clustering with intactness-aware similarity. Pattern Recognit 88:50–63

122. Guo J, Yin W, Sun Y, Hu Y (2019) Multi-view subspace clustering with block diagonal representation. IEEE Access 7:84829–84838

123. Guo J, Sun Y, Gao J, Hu Y, Yin B (2020) Low rank representation on product grassmann manifolds for multi-view subspace clustering. In: ICPR, pp 907–914

124. Sun M, Zhang P, Wang S, Zhou S, Tu W, Liu X, Zhu E, Wang C (2021) Scalable multi-view subspace clustering with unified anchors. In: MM, pp 3528–3536

125. Zhang C, Fu H, Liu S, Liu G, Cao X (2015) Low-rank tensor constrained multiview subspace clustering. In: ICCV

126. Feng L, Cai L, Liu Y, Liu S (2017) Multi-view spectral clustering via robust local subspace learning. Soft Comput 21(8):1937–1948

127. Xie Y, Tao D, Zhang W, Liu Y, Zhang L, Qu Y (2018) On unifying multi-view self-representations for clustering by tensor multi-rank minimization. Int J Comput Vis 126(11):1157–1179

128. Gao Q, Xia W, Wan Z, Xie D, Zhang P (2020) Tensor-svd based graph learning for multi-view subspace clustering. In: AAAI, pp 3930–3937

129. Liu Y, Zhang X, Tang G, Wang D (2019) Multi-view subspace clustering based on tensor schatten-p norm. In: IEEE BigData, pp 5048–5055

130. Sun X, Wang Y, Zhang X (2020) Multi-view subspace clustering via non-convex tensor rank minimization. In: ICME, pp 1–6

131. Chen Y, Xiao X, Zhou Y (2020) Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix. Pattern Recognit 106:107441

132. Chen Y, Wang S, Peng C, Hua Z, Zhou Y (2021) Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering. IEEE Trans Image Process 30:4022–4035

133. Qian M, Zhai C (2014) Unsupervised feature selection for multi-view clustering on text-image web news data. In: CIKM, pp 1963–1966

134. Xu Y, Wang C, Lai J (2016) Weighted multi-view clustering with feature selection. Pattern Recognit 53:25–35

135. Yin Q, Zhang J, Wu S, Li H (2019) Multi-view clustering via joint feature selection and partially constrained cluster label learning. Pattern Recognit 93:380–391

136. Wang C, Lai J, Yu PS (2016) Multi-view clustering based on belief propagation. IEEE Trans Knowl Data Eng 28(4):1007–1021

137. Zhang Y, Chung F, Wang S (2019) A multiview and multiexemplar fuzzy clustering approach: theoretical analysis and experimental studies. IEEE Trans Fuzzy Syst 27(8):1543–1557

138. Zhang Z, Liu L, Shen F, Shen HT, Shao L (2019) Binary multi-view clustering. IEEE Trans Pattern Anal Mach Intell 41(7):1774–1782

139. Yu X, Liu H, Wu Y, Zhang C (2021) Fine-grained similarity fusion for multi-view spectral clustering. Inf Sci 568:350–368

140. Xie J, Girshick RB, Farhadi A (2016) Unsupervised deep embedding for clustering analysis. ICML 48:478–487

141. Abavisani M, Patel VM (2018) Deep multimodal subspace clustering networks. IEEE J Sel Top Signal Process 12(6):1601–1614

142. Bai R, Huang R, Chen Y, Qin Y (2021) Deep multi-view document clustering with enhanced semantic embedding. Inf Sci 564:273–287

143. Xu J, Ren Y, Li G, Pan L, Zhu C, Xu Z (2021) Deep embedded multi-view clustering with collaborative training. Inf Sci 573:279–290

144. Yang X, Deng C, Dang Z, Tao D (2021) Deep multiview collaborative clustering. IEEE Trans Neural Networks Learn Syst

145. Wang Q, Cheng J, Gao Q, Zhao G, Jiao L (2021) Deep multi-view subspace clustering with unified and discriminative learning. IEEE Trans Multim 23:3483–3493

146. Zheng Q, Zhu J, Ma Y, Li Z, Tian Z (2021) Multi-view subspace clustering networks with local and global graph information. Neurocomputing 449:15–23

147. Lu R, Liu J, Zuo X (2021) Attentive multi-view deep subspace clustering net. Neurocomputing 435:186–196

148. Chen L, Yang Y, Wang J, Xu W, Yuille AL (2016) Attention to scale: scale-aware semantic image segmentation. In: CVPR, pp 3640–3649

149. Chang B, Krüger U, Kustra R, Zhang J (2013) Canonical correlation analysis based on Hilbert-Schmidt independence criterion and centered kernel target alignment. ICML 28:316–324

150. Zhu P, Hui B, Zhang C, Du D, Wen L, Hu Q (2019) Multi-view deep subspace clustering networks. CoRR arXiv:1908.01978

151. Cui B, Yu H, Zong L, Cheng Z (2021) Self-Guided deep multi-view subspace clustering network. In: ICME, pp 1–6

152. Ke G, Hong Z, Yu W, Zhang X, Liu Z (2022) Efficient multi-view clustering networks. Appl Intell 1–17

153. Zhang Y, Huang Q, Zhang B, He S, Dan T, Peng H, Cai H (2021) Deep multiview clustering via iteratively Self-Supervised universal and specific space learning. IEEE Trans Cybern

154. Du G, Zhou L, Yang Y, Lü K, Wang L (2021) Deep multiple Auto-Encoder-Based multi-view clustering. Data Sci Eng 6(3):323–338

155. Trosten DJ, Løkse S, Jenssen R, Kampffmeyer M (2021) Reconsidering representation alignment for multi-view clustering. In: CVPR, pp 1255–1265

156. Wang S, Xiao S, Zhu W, Guo Y (2022) Multi-view fuzzy clustering of deep random walk and sparse low-rank embedding. Inf Sci 586:224–238

157. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G (2009) The graph neural network model. IEEE Trans Neural Netw Learn Syst 20(1):61–80

158. Kipf TN, Welling M (2016) Semi-Supervised classification with graph convolutional networks. CoRR arXiv:1609.02907

159. Fan S, Wang X, Shi C, Lu E, Lin K, Wang B (2020) One2Multi graph autoencoder for multi-view graph clustering. In: WWW, pp 3070–3076

160. Cheng J, Wang Q, Tao Z, Xie D, Gao Q (2020) Multi-view attribute graph convolution networks for clustering. In: IJCAI, pp 2973–2979

161. Zhang X, Yang Y, Zhai D, Li T, Chu J, Wang H (2021) Local2Global: unsupervised multi-view deep graph representation learning with nearest neighbor constraint. Knowl Based Syst 231:107439

162. Sarkar A, Mehta N, Rai P (2020) Graph representation learning via ladder gamma variational autoencoders. In: AAAI, pp 5604–5611

163. Cai E, Huang J, Huang B, Xu S, Zhu J (2021) GRAE: graph recurrent autoencoder for multi-view graph clustering. In: ACAI, pp 72–1729

164. Xia W, Wang S, Yang M, Gao Q, Han J, Gao X (2022) Multi-view graph embedding clustering network: Joint self-supervision and block diagonal representation. Neural Netw 145:1–9

165. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville AC, Bengio Y (2014) Generative adversarial nets. In: NIPS, pp 2672–2680

166. Li Z, Wang Q, Tao Z, Gao Q, Yang Z (2019) Deep adversarial multi-view clustering network. In: IJCAI, pp 2952–2958

167. Wang Q, Tao Z, Xia W, Gao Q, Cao X, Jiao L (2022) Adversarial multiview clustering networks with adaptive fusion. IEEE Trans Neural Netw Learn Syst 1–13

168. Zhou R, Shen Y (2020) End-to-End adversarial-attention network for multi-modal clustering. In: CVPR, pp 14607–14616

169. Li Y, Liao H (2021) Multi-view clustering via adversarial view embedding and adaptive view fusion. Appl Intell 51(3):1201–1212

170. Huang Z, Zhou JT, Peng X, Zhang C, Zhu H, Lv J (2019) Multi-view spectral clustering network. In: IJCAI, pp 2563–2569

171. Ke G, Hong Z, Zeng Z, Liu Z, Sun Y, Xie Y (2021) CONAN: contrastive fusion networks for multi-view clustering. In: IEEE BigData, pp 653–660

172. Xin B, Zeng S, Wang X (2021) Self-Supervised deep correlational multi-view clustering. In: IJCNN, pp 1–8

173. Burgess CP, Higgins I, Pal A, Matthey L, Watters N, Desjardins G, Lerchner A (2018) Understanding disentangling in β-vae. CoRR arXiv:1804.03599

174. Xu J, Ren Y, Tang H, Pu X, Zhu X, Zeng M, He L (2021) Multi-VAE: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering. In: ICCV, pp 9214–9223

175. Xie Y, Lin B, Qu Y, Li C, Zhang W, Ma L, Wen Y, Tao D (2021) Joint deep Multi-View learning for image clustering. IEEE Trans Knowl Data Eng 33(11):3594–3606

176. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987

177. Lades M, Vorbruggen JC, Buhmann J, Lange J, von der Malsburg C, Wurtz RP, Konen W (1993) Distortion invariant object recognition in the dynamic link architecture. IEEE Trans Comput 42(3):300–311

178. Shen D, Ip HH (1999) Discriminative wavelet shape descriptors for recognition of 2-d patterns. Pattern Recogn 32(2):151–165

179. Wu J, Rehg JM (2011) Centrist: a visual descriptor for scene categorization. IEEE Trans Pattern Anal Mach Intell 33(8):1489–1501

180. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: CVPR, pp 886–893

181. Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis 42(3):145–175

182. Kumar G, Bhatia PK (2014) A detailed review of feature extraction in image processing systems. In: ICACCT, pp 5–12

183. Fukunaga K, Koontz WL (1970) Application of the Karhunen–Loeve expansion to feature selection and ordering. IEEE Trans Comput 100(4):311–318

184. Khotanzad A, Hong YH (1990) Invariant image recognition by Zernike moments. IEEE Trans Pattern Anal Mach Intell 12(5):489–497

185. Wen J, Zhang Z, Xu Y, Zhang B, Fei L, Liu H (2019) Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In: AAAI, pp 5393–5400

186. Wen J, Zhang Z, Zhang Z, Zhu L, Fei L, Zhang B, Xu Y (2021) Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In: AAAI, pp 10273–10281

187. Yu H, Tang J, Wang G, Gao X (2021) A novel multi-view clustering method for unknown mapping relationships between cross-view samples. In: KDD, pp 2075–2083

188. Tekumalla LS, Rajan V, Bhattacharyya C (2017) Vine copulas for mixed data: multi-view clustering for mixed data beyond meta-gaussian dependencies. Mach Learn 106(9–10):1331–1357

189. Wang M, Allen GI (2021) Integrative generalized convex clustering optimization and feature selection for mixed multi-view data. J Mach Learn Res 22:55–15573