

Article

Adaptive Weighted Graph Fusion Incomplete Multi-View Subspace Clustering

Pei Zhang , Siwei Wang, Jingtao Hu, Zhen Cheng, Xifeng Guo, En Zhu * and Zhiping Cai

School of Computer, National University of Defense Technology, Changsha 410073, China; zhangpei@nudt.edu.cn (P.Z.); wangsiwei13@nudt.edu.cn (S.W.); hujingtao17@nudt.edu.cn (J.H.); chengzhen16@nudt.edu.cn (Z.C.); guoxifeng13@nudt.edu.cn (X.G.); zpc@nudt.edu.cn (Z.C.)

* Correspondence: enzhu@nudt.edu.cn

Received: 24 August 2020; Accepted: 6 October 2020; Published: 10 October 2020

Abstract: With the enormous amount of multi-source data produced by various sensors and feature extraction approaches, multi-view clustering (MVC) has attracted developing research attention and is widely exploited in data analysis. Most of the existing multi-view clustering methods hold on the assumption that all of the views are complete. However, in many real scenarios, multi-view data are often incomplete for many reasons, e.g., hardware failure or incomplete data collection. In this paper, we propose an adaptive weighted graph fusion incomplete multi-view subspace clustering (AWGF-IMSC) method to solve the incomplete multi-view clustering problem. Firstly, to eliminate the noise existing in the original space, we transform complete original data into latent representations which contribute to better graph construction for each view. Then, we incorporate feature extraction and incomplete graph fusion into a unified framework, whereas two processes can negotiate with each other, serving for graph learning tasks. A sparse regularization is imposed on the complete graph to make it more robust to the view-inconsistency. Besides, the importance of different views is automatically learned, further guiding the construction of the complete graph. An effective iterative algorithm is proposed to solve the resulting optimization problem with convergence. Compared with the existing state-of-the-art methods, the experiment results on several real-world datasets demonstrate the effectiveness and advancement of our proposed method.

Keywords: multi-feature; incomplete multi-view clustering; subspace learning; graph fusion

1. Introduction

Traditional clustering methods [1–4] usually use a single view to measure the similarity of samples. With the rapid progress of data collection, individual features are not enough to describe data points. Multiple views usually contain supplementary information, which may be beneficial to explore the basic structure of the data. With the development of information technology, data mining and other technologies, many datasets in the real-world can be presented from different perspectives, called multi-view data. For example, the same text can be expressed in various languages. In biometric recognition scope, faces, fingerprints, palm prints and iris could form the different views of multi-view data. In the field of medical diagnosis, different examinations of patients can be regarded as different views. Multi-view data could provide sufficient information than the traditional single feature representation in revealing the underlying clustering structure. Furthermore, distinct views contain specific information of intra-view and complementary information of inter-view, which are negotiated with each other to boost the performance of clustering [5–14].

Based on different mechanisms, we can divide the existing multi-view clustering methods into four categories. The first category methods refer to multi-kernel clustering. These methods usually combine multiple pre-defined kernels to reach optimal clustering results [12,15–17]. The second kind

of approach is co-training and co-regularized [18,19]. They iteratively learn multiple clustering results that can provide predicted clustering indices for the unlabeled data from different views. In this way, the clustering results are forced to be consistent across views. The third strategy collaboratively transforms the multi-view information into a compact common binary code space. Then, the clustering process is measured in the Hamming space and enjoys superior algorithm acceleration [20,21]. The last mechanism is the subspace-based multi-view clustering method. It assumes that high-dimensional data points are drawn from various low-dimensional subspaces, and each cluster can be drawn from one of the subspaces [22–25]. The essential idea is to find several low-dimensional representations embedded in latent spaces and finally attain a united representation for downstream clustering tasks [26]. Besides, aiming at finding a shared low-dimensional latent representation via matrix decomposition, the non-negative matrix factorization (NMF) [27]-based multi-view clustering methods [28–31] can also be seen as a branch of the subspace-based multi-view clustering method.

Although the algorithms mentioned above have achieved great success in different scenarios, these traditional multi-view clustering algorithms cannot effectively deal with multi-view data with incomplete features. Therefore, the incomplete multi-view clustering algorithms [32–34] have attracted extensive attention. To the best of our knowledge, existing incomplete multi-view clustering algorithms can be classified into two categories: non-negative matrix factorization based methods and graph-based methods. The NMF-based methods aim at directly obtaining a common low-dimensional representation through non-negative matrix decomposition. Most of them take the strategy of combining view-specific and common representations into a unified one [35–37]. Another representative approach of the NMF-based method is to fill the missing data with average feature values and then use the weighted non-negative matrix factorization to reduce the impact of the missing samples [38]. These NMF-based methods can directly obtain a consistent representation with incomplete samples. However, it is limited to the following two points: (1) when the number of views is more than two, the common parts of views will be significantly reduced and cannot be learned a shared representation between views; (2) NMF-based methods usually neglect the intrinsic structure of data, resulting in an uncompact representation.

The graph-based incomplete multi-view clustering algorithms are more effective in exploring the geometric structure of data than NMF-based methods. The construction of the graph is essential for the success of clustering. However, it is impossible to construct a complete graph connecting all samples due to the lack of partial samples in incomplete multi-view clustering. To cover this problem, Gao et al. [39] first fill the missing parts and then learn graphs and representations. Zhao et al. [36] utilize NMF to obtain consistent representations to guide the generation of graphs with local structures. However, when the missing rate is high, the filling strategy will dominate the learning of the representation, resulting in the filled samples being connected with each other. Moreover, information fusion refers to fusing multiple sources to achieve consistency. In this stage, multiple views are treated equally, which is unreasonable in real applications.

To address the above issues, we propose a novel incomplete multi-view clustering method, constructing the graphs between instances in the latent embedding subspace. In this manner, we can deal with multi-view data with any number of views. Furthermore, an adaptive weighted mechanism is induced to fuse the graphs with local-structure into a complete graph.

In this manner, we establish a relation between missing and unmissing samples. An additional sparse regularization term is imposed on the consensus complete graph to eliminate the adverse effects of inconsistency between views and noise or outliers from each view. Specifically, the framework of this paper is illustrated in Figure 1.

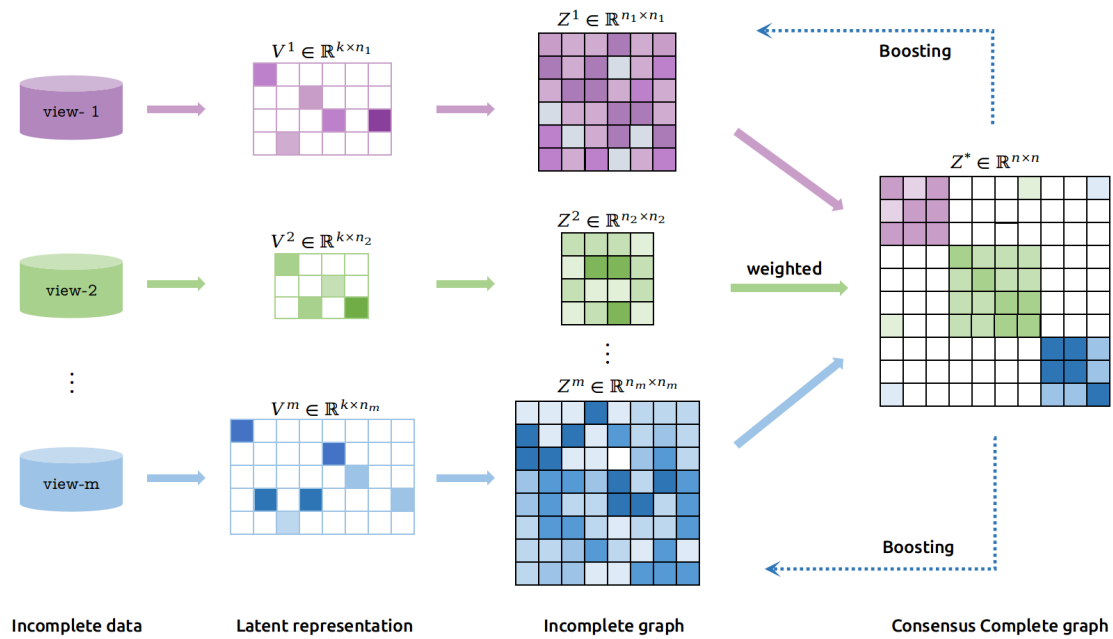


Figure 1. Framework of the proposed adaptive weighted graph fusion incomplete multi-view subspace clustering (AWGF-IMSC). It is a novel incomplete multi-view clustering method to fuse the local-structure contained graph with adaptive view-importance learning. Incomplete graphs of different scales are fused into a complete graph with automatically learning weights. In addition, the constructed complete graph will further guide the learning process of incomplete graphs and latent representations.

Compared with existing methods, the proposed adaptive weighted graph fusion incomplete multi-view subspace clustering (AWGF-IMSC) algorithm has the following contributions:

- It induces the similarity graph fusion after obtaining latent spaces to extract the local structure of inner views. By virtue of it, noise existing in the original space can be eliminated in latent space and contribute to better graph construction.
- It incorporates relations between missing samples and complete samples into the complete graph. The sparse constraint imposed on the complete graph improves the view-inconsistency and reduces the disagreements between views, making the proposed method more robust in most cases.
- The importance of each view is automatically learned and adaptively optimized during the optimization. Consequently, the important view has strong guidance in the learning process. Moreover, there is no limitation to the number of views in our approach. The proposed method is applicable to any multi-view datasets.

The rest of the paper is organized as follows. The next Section 2 denotes the notations and symbols used in this paper. Section 3 introduces methods mostly related to our work. The proposed algorithm and its optimization process are formulated in Section 4. Besides, we also give the analysis of convergence and complexity of the proposed algorithm in this part. Extensive experiment results and analysis are shown in Section 5, before conclusion and perspectives.

2. Notation

For clarity, we give the notation used throughout the paper at the beginning. We use bold letters to represent matrices and vectors. For matrix \mathbf{A} , $\mathbf{A}_{:,j}$ and $A_{i,j}$ represent its j -th column and (i, j) element, respectively. \mathbf{A}^\top , $\text{Tr}(\mathbf{A})$ and \mathbf{A}^{-1} denote the transpose, trace and the inverse operations on matrix \mathbf{A} , respectively. $\|\cdot\|_F$ denotes the Frobenius norm. The $\ell_{2,1}$ norm is denoted as $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^t \mathbf{X}_{i,j}^2} = \sum_{i=1}^n \|\mathbf{X}_{i,:}\|_2$. Moreover, operator \mathbf{A}^+ turns the negative elements in matrix \mathbf{A} to 0 while maintaining

the non-negative elements, and vice versa. For multi-view datasets $\mathbf{A} = \{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(m)}\}$, the superscript (i) represents the i -th view. In individual $\mathbf{A}^{(i)} \in \mathbb{R}^{d_i \times N}$, each column indicates an instance. N is the number of the samples and d_i represents the feature dimension of corresponding i -th view.

3. Related Work

In this section, we will present the work most relevant to the proposed method, i.e., semi-non-negative matrix factorization and subspace learning.

3.1. Semi-Non-Negative Matrix Factorization for Single View

Non-negative matrix factorization (NMF) is a significant branch in the field of matrix factorization. NMF aims at finding two non-negative matrix $\mathbf{U} \in \mathbb{R}_+^{d \times K}$ and $\mathbf{V} \in \mathbb{R}_+^{K \times N}$ to roughly approximate the original data matrix, i.e., $\mathbf{X} \approx \mathbf{UV}$. Since many real-world datasets are usually high-dimensional, the NMF methods have been widely applied in image analysis [40], data mining, speech denoising [41] and population genetics, etc. The semi-NMF [42] is an extension of traditional NMF, which only requires the coefficient matrix to be non-negative. Specifically, given the data matrix $\mathbf{X} \in \mathbb{R}^{d \times N}$, the semi-NMF utilizes the base matrix $\mathbf{U} \in \mathbb{R}^{d \times k}$ and the non-negative coefficient matrix $\mathbf{V} \in \mathbb{R}^{k \times N}$ to approximate the matrix \mathbf{X} :

$$\min_{\mathbf{U}, \mathbf{V}} \|\mathbf{X} - \mathbf{UV}\|_F^2 \quad \text{s.t. } \mathbf{V} \geq 0. \quad (1)$$

Ding et al. [42] further propose an iterative optimization algorithm to find the local optimal solution. The updating strategy can be concluded as follows:

With \mathbf{V} being fixed, \mathbf{U} can be updated by $\mathbf{U} = \mathbf{XV}^T(\mathbf{VV}^T)^{-1}$.

with \mathbf{U} being fixed, \mathbf{V} can be updated by $V_{i,j} \leftarrow V_{i,j} \sqrt{\frac{(\mathbf{X}^T \mathbf{U})_{i,j}^+ + [\mathbf{V}^T (\mathbf{U}^T \mathbf{U})^-]_{i,j}}{(\mathbf{X}^T \mathbf{U})_{i,j}^- + [\mathbf{V}^T (\mathbf{U}^T \mathbf{U})^+]_{i,j}}}$.

The positive and negative elements of matrix \mathbf{M} are denoted as $M_{i,j}^+$ and $M_{i,j}^-$. And they hold on the property $M_{i,j} = M_{i,j}^+ - M_{i,j}^-$.

NMF and semi-NMF methods are also employed universally in multi-view clustering. Many of the multi-view clustering (MVC) methods utilize NMF to reduce dimension on each view or directly reach a consistent latent presentation [28,43]. Especially in the incomplete multi-view scenario, NMF and semi-NMF play significant roles in achieving a consistent representation from different incomplete views. Li et al. [35] learn a shared representation for the paired instances and view-specific representations for unpaired instances via NMF. The complete latent representation can be attained by combining shared and view-specific representations. The method in [44] utilizes weighted semi-NMF to reach a consensus representation. Then, the $\ell_{2,1}$ norm regularized regression is imposed to align the different basis matrices. Although these NMF-based methods could learn a consensus representation from the incomplete views, the number of views and the absence of local structure limit their performance.

3.2. Subspace Clustering

Subspace clustering is an extension of the traditional clustering method which aims at grouping data in different subspaces [45,46]. The self-representation property [47] of subspace clustering aims to represent data points by the linear combinations of themselves. The formulation can be expressed as:

$$\begin{aligned} \min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{XZ}\|_F^2 + \beta \mathbf{R}(\mathbf{Z}) \\ \text{s.t. } 0 \leq Z_{i,j} \leq 1, \mathbf{Z}^T \mathbf{1} = \mathbf{1}, \end{aligned} \quad (2)$$

where $\mathbf{X} \in \mathbb{R}^{d \times N}$ is the original data, \mathbf{Z} is the self-representation coefficient matrix, with each column being a new representation for corresponding data point. $\beta > 0$ is a trade-off parameter. Since \mathbf{Z}

reflects the correlations among samples, it can be regarded as a graph and then we can perform spectral clustering algorithm on it to get the final clustering result.

3.3. Incomplete Multi-View Spectral Clustering with Adaptive Graph Learning (IMSC-AGL)

In paper [48], a novel graph-based multi-view clustering method is proposed to deal with incomplete multi-view scenarios termed incomplete multi-view spectral clustering with adaptive graph learning (IMSC-AGL). IMSC-AGL optimizes the shared graph from the low-dimensional representations individually formed by each view. Moreover, a nuclear-norm constraint is introduced to ensure the low-rank property of the ideal graph. The mathematical formulation can be written as,

$$\begin{aligned} \min_{\mathbf{Z}^{(v)}, \mathbf{E}^{(v)}, \mathbf{F}^{(v)}, \mathbf{M}} \quad & \sum_v \left(\left\| \mathbf{Z}^{(v)} \right\|_* - \lambda_1 \text{Tr} \left(\mathbf{F}^{(v)} \mathbf{F}^{(v)T} \mathbf{M} \mathbf{M}^T \right) \right) \\ & + \sum_v \lambda_2 \text{Tr} \left(\mathbf{F}^{(v)T} \mathbf{G}^{(v)T} \mathbf{L}^{(v)} \mathbf{G}^{(v)} \mathbf{F}^{(v)} \right) + \sum_v \lambda_3 \left\| \mathbf{E}^{(v)} \right\|_1 \\ \text{s.t.} \quad & \mathbf{Y}^{(v)} = \mathbf{Y}^{(v)} \mathbf{Z}^{(v)} + \mathbf{E}^{(v)}, \mathbf{Z}^{(v)} \mathbf{1} = \mathbf{1}, \\ & 0 \leq \mathbf{Z}^{(v)} \leq \mathbf{1}, \mathbf{Z}_{ii}^{(v)} = 0, \mathbf{F}^{(v)T} \mathbf{F}^{(v)} = \mathbf{I}, \mathbf{M}^T \mathbf{M} = \mathbf{I}, \end{aligned} \quad (3)$$

where \mathbf{Y}^v represents the complete samples in v -th view. $\mathbf{Z}^{(v)}$ denotes the respective v -th view's graph. $\mathbf{F}^{(v)}$ represents the clustering indicator matrix with proper size. Moreover, \mathbf{M} refers to the final shared clustering indicator matrix. Although IMSC-AGL achieves considerable performance in various applications, it can still be improved from the number of hyper-parameters and considering to fuse multiple information in a weighted manner.

4. Method

4.1. Adaptive Weighted Graph Fusion Incomplete Multi-View Subspace Clustering

In this section, we present our adaptive graph fusion incomplete multi-view subspace clustering method (AWGF-IMSC) in detail and give a unified objective function.

For incomplete multi-view data, we remove the incomplete instances and reform as $\mathbf{X}^{(i)} \in \mathbb{R}^{d_i \times n_i}$, where d_i and n_i represent the feature dimension and the numbers of visible samples of i -th view, respectively. We assume that semi-NMF factorizes the input data $\mathbf{X}^{(i)} \in \mathbb{R}^{d_i \times n_i}$ into base matrix $\mathbf{U}^{(i)} \in \mathbb{R}^{d_i \times k}$ and coefficient matrix $\mathbf{V}^{(i)} \in \mathbb{R}^{k \times n_i}$. k is the dimension of target space and is commonly set to the number of the clusters of $\mathbf{X}^{(i)}$. Considering that the missing samples differ in each view, we learn latent representations of the corresponding visible samples in each view. Therefore, the semi-NMF for individual view can be formulated as:

$$\min_{\mathbf{U}^{(i)}, \mathbf{V}^{(i)}} \sum_{i=1}^m \left\| \mathbf{X}^{(i)} - \mathbf{U}^{(i)} \mathbf{V}^{(i)} \right\|_F^2 \quad \text{s.t.} \quad \mathbf{V}^{(i)} \geq 0. \quad (4)$$

To further exploit the intra-view similarity structure and the underlying subspace structure, we utilize the self-representation property [47] on the $k \times n_i$ dimensional latent representation $\mathbf{V}^{(i)}$ to construct the graph. Thus, we can obtain the different graphs $\mathbf{Z}^{(i)}$ of individual views by solving the following problem:

$$\begin{aligned} \min_{\mathbf{V}^{(i)}, \mathbf{Z}^{(i)}} \quad & \sum_{i=1}^m \left\| \mathbf{V}^{(i)} - \mathbf{V}^{(i)} \mathbf{Z}^{(i)} \right\|_F^2 \\ \text{s.t.} \quad & \mathbf{V}^{(i)} \geq 0, 0 \leq \mathbf{Z}_{j,k}^{(i)} \leq 1, \mathbf{Z}^{(i)T} \mathbf{1} = \mathbf{1}, \end{aligned} \quad (5)$$

where the constraint $0 \leq \mathbf{Z}_{j,k}^{(i)} \leq 1$ and $\mathbf{Z}^{(i)T} \mathbf{1} = \mathbf{1}$ guarantee a good probabilistic explanation for $\mathbf{Z}^{(i)}$. After obtaining the graphs on each view, the natural idea is to integrate the multiple incomplete information into a complete one. In order to establish the correspondence between the incomplete

and complete graphs, we denote the index matrix $\mathbf{O}^{(i)}$. The index matrix $\mathbf{O}^{(i)} \in \mathbb{R}^{n_i \times N}$ can extract the visible instances of view i from the complete graph. To be specific, the matrix $\mathbf{O}^{(i)}$ is defined as:

$$\mathbf{O}_{j,k}^{(i)} = \begin{cases} 1, & \text{if } x_j^{(i)} \text{ is the } k\text{-th sample in complete dataset.} \\ 0, & \text{otherwise.} \end{cases}$$

Through the index matrix, we can achieve the transformation between complete and incomplete graphs: $\mathbf{Z}^{(i)} = \mathbf{O}^{(i)} \mathbf{Z}^* \mathbf{O}^{(i)\top}$ or $\mathbf{O}^{(i)\top} \mathbf{Z}^{(i)} \mathbf{O}^{(i)} = \hat{\mathbf{Z}}^*$. In the second condition, $\mathbf{O}^{(i)}$ expands the graph $\mathbf{Z}^{(i)}$ into $\hat{\mathbf{Z}}^*$, where $\hat{\mathbf{Z}}^*$ has the same size with \mathbf{Z}^* , but the irrelevant items to view i are zero.

Owing to the size of the graph and the similarity magnitude differing among views, it is unreasonable to directly add up the multiple graphs. Consequently, we aim to integrate the multiple information into a completed graph with adaptive learning weights $\{\alpha_i\}_{i=1}^m$. With the help of the index matrix, relevant elements can be extracted from \mathbf{Z}^* . Then, we can adaptively fuse $\{\mathbf{Z}^{(i)}\}_{i=1}^m$ into a complete graph with auto-learning weights, as illustrated in Equation (6).

$$\min_{\alpha_i} \sum_{i=1}^m \left\| \alpha_i \mathbf{Z}^{(i)} - \mathbf{O}^{(i)} \mathbf{Z}^* \mathbf{O}^{(i)\top} \right\|_F^2 \quad \text{s.t. } \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1, \quad (6)$$

where α_i is the weight for i -th view. It is automatically learned and optimized to illustrate the importance of i -th view. In this manner, the complete graph is learned by a weighted combination of incomplete graphs. Besides, with the fusion of beneficial information, the inconsistencies between different views, noise and outliers in individual view are also integrated into the complete graph. Considering that, an additional sparse constraint is added on \mathbf{Z}^* . Therefore, integrating the above parts into an unified objective function, we have our optimization goal as:

$$\begin{aligned} \min_{\substack{\mathbf{U}^{(i)}, \mathbf{V}^{(i)}, \\ \alpha_i, \mathbf{Z}^{(i)}, \mathbf{Z}^*}} \sum_{i=1}^m \left\| \mathbf{X}^{(i)} - \mathbf{U}^{(i)} \mathbf{V}^{(i)} \right\|_F^2 + \left\| \mathbf{V}^{(i)} - \mathbf{V}^{(i)} \mathbf{Z}^{(i)} \right\|_F^2 + \lambda_1 \left\| \alpha_i \mathbf{Z}^{(i)} - \mathbf{O}^{(i)} \mathbf{Z}^* \mathbf{O}^{(i)\top} \right\|_F^2 + \lambda_2 \left\| \mathbf{Z}^* \right\|_1 \\ \text{s.t. } \mathbf{V}^{(i)} \geq 0, 0 \leq \mathbf{Z}_{j,k}^{(i)} \leq 1, \mathbf{Z}^{(i)\top} \mathbf{1} = \mathbf{1}, \sum_{i=1}^m \alpha_i = 1, \alpha_i \geq 0, \end{aligned} \quad (7)$$

λ_1 and λ_2 are non-negative trade-off parameters. In the proposed framework, we have four terms: using semi-NMF to obtain latent representation, conducting graph construction with self-representation, adaptive graph fusion and sparse regularizer. Finally, we get a full-size graph \mathbf{Z}^* incorporating all the sample information in the latent subspace.

4.2. Optimization Algorithm for AWGF-IMSC

The constraint problem in Equation (7) is not jointly convex with regard to all the variables. In this section, we propose an alternating iterative algorithm to solve this optimization problem.

4.2.1. Update $\mathbf{U}^{(i)}$

With $\mathbf{V}^{(i)}$, $\mathbf{Z}^{(i)}$, α_i and \mathbf{Z}^* fixed, for each $\mathbf{U}^{(i)}$, we need to solve the following problem,

$$\min_{\mathbf{U}^{(i)}} \left\| \mathbf{X}^{(i)} - \mathbf{U}^{(i)} \mathbf{V}^{(i)} \right\|_F^2 \quad (8)$$

Each of $\mathbf{U}^{(i)}$ can be solved separately since views are independent from each other. Therefore, the optimization problem that we minimize can be rewritten as:

$$\mathcal{L}(\mathbf{U}^{(i)}) = \text{Tr}(\mathbf{X}^{(i)\top} \mathbf{X}^{(i)} - 2\mathbf{V}^{(i)\top} \mathbf{U}^{(i)\top} \mathbf{X}^{(i)} + \mathbf{V}^{(i)\top} \mathbf{U}^{(i)\top} \mathbf{U}^{(i)} \mathbf{V}^{(i)}) \quad (9)$$

The solution for $\mathbf{U}^{(i)}$ can be easily obtained by setting the derivation w.r.t. $\mathbf{U}^{(i)}$ to zero.

$$\frac{\partial \mathcal{L}(\mathbf{U}^{(i)})}{\partial \mathbf{U}^{(i)}} = -2\mathbf{X}^{(i)}\mathbf{V}^{(i)\top} + 2\mathbf{U}^{(i)}\mathbf{V}^{(i)}\mathbf{V}^{(i)\top} \quad (10)$$

Then, we can get the optimal closed-form solution:

$$\mathbf{U}^{(i)} = \mathbf{X}^{(i)}\mathbf{V}^{(i)\top}(\mathbf{V}^{(i)}\mathbf{V}^{(i)\top})^{-1} \quad (11)$$

4.2.2. Update $\mathbf{V}^{(i)}$

Fixing $\mathbf{U}^{(i)}$, $\mathbf{Z}^{(i)}$, α_i and \mathbf{Z}^* , the minimum problem for optimizing $\mathbf{V}^{(i)}$ can be simplified as:

$$\mathcal{L}(\mathbf{V}^{(i)}) = \sum_{i=1}^m \left\| \mathbf{X}^{(i)} - \mathbf{U}^{(i)}\mathbf{V}^{(i)} \right\|_F^2 + \left\| \mathbf{V}^{(i)} - \mathbf{V}^{(i)}\mathbf{Z}^{(i)} \right\|_F^2 \quad \text{s.t. } \mathbf{V}^{(i)} \geq 0, \quad (12)$$

We can update $\mathbf{V}^{(i)}$ in Equation (12) referring to the update strategy in semi-NMF, since the semi-NMF and subspace learning processes are isolated from each other. The partial derivation of $\mathcal{L}(\mathbf{V}^{(i)})$ with respect to $\mathbf{V}^{(i)}$ can be obtained as:

$$\frac{\partial \mathcal{L}(\mathbf{V}^{(i)})}{\partial \mathbf{V}^{(i)}} = \mathbf{V}^{(i)}(\mathbf{I} - \mathbf{Z}^{(i)} - \mathbf{Z}^{(i)\top} + \mathbf{Z}^{(i)}\mathbf{Z}^{(i)\top}) + \mathbf{U}^{(i)\top}\mathbf{U}^{(i)}\mathbf{V}^{(i)} - \mathbf{U}^{(i)\top}\mathbf{X}^{(i)}. \quad (13)$$

According to the optimization of semi-NMF and the KKT condition, we can get

$$(\mathbf{V}^{(i)}\mathbf{P}^{(i)} + \mathbf{U}^{(i)\top}\mathbf{U}^{(i)}\mathbf{V}^{(i)} - \mathbf{U}^{(i)\top}\mathbf{X}^{(i)})_{j,k}\mathbf{V}_{j,k}^{(i)} = 0,$$

where $\mathbf{P}^{(i)} = \mathbf{I} - \mathbf{Z}^{(i)} - \mathbf{Z}^{(i)\top} + \mathbf{Z}^{(i)}\mathbf{Z}^{(i)\top}$. Note that $\mathbf{U}^{(i)\top}\mathbf{U}^{(i)} = (\mathbf{U}^{(i)\top}\mathbf{U}^{(i)})^+ - (\mathbf{U}^{(i)\top}\mathbf{U}^{(i)})^-$. Based on this, we can achieve the updating rule for $\mathbf{V}^{(i)}$:

$$V_{j,k}^{(i)} \leftarrow V_{j,k}^{(i)} \sqrt{\frac{(\mathbf{V}^{(i)}(\mathbf{P}^{(i)})^-)_{j,k} + [(\mathbf{U}^{(i)\top}\mathbf{U}^{(i)})^- \mathbf{V}^{(i)}]_{j,k} + (\mathbf{U}^{(i)\top}\mathbf{X}^{(i)})_{j,k}^+}{(\mathbf{V}^{(i)}(\mathbf{P}^{(i)})^+)_{j,k} + [(\mathbf{U}^{(i)\top}\mathbf{U}^{(i)})^+ \mathbf{V}^{(i)}]_{j,k} + (\mathbf{U}^{(i)\top}\mathbf{X}^{(i)})_{j,k}^-}}. \quad (14)$$

4.2.3. Update $\mathbf{Z}^{(i)}$

When $\mathbf{U}^{(i)}$, $\mathbf{V}^{(i)}$, α_i and \mathbf{Z}^* fixed, the optimization for $\mathbf{Z}^{(i)}$ can be simplified as:

$$\min_{\mathbf{Z}^{(i)}} \sum_{i=1}^m \left\| \mathbf{V}^{(i)} - \mathbf{V}^{(i)}\mathbf{Z}^{(i)} \right\|_F^2 + \lambda_1 \left\| \alpha_i \mathbf{Z}^{(i)} - \mathbf{O}^{(i)}\mathbf{Z}^*\mathbf{O}^{(i)\top} \right\|_F^2 \quad (15)$$

$$\text{s.t. } 0 \leq \mathbf{Z}_{j,k}^{(i)} \leq 1, \mathbf{Z}^{(i)\top} \mathbf{1} = \mathbf{1}.$$

Denoting $\mathbf{H}^{(i)} = \mathbf{O}^{(i)}\mathbf{Z}^*\mathbf{O}^{(i)\top}$, we can obtain the following equivalent question

$$\min_{\mathbf{Z}^{(i)}} \sum_{i=1}^m \text{Tr}(\mathbf{Z}^{(i)\top}(\mathbf{V}^{(i)\top}\mathbf{V}^{(i)} + \lambda_1 \alpha_i^2 \mathbf{I})\mathbf{Z}^{(i)} - 2\mathbf{Z}^{(i)\top}(\mathbf{V}^{(i)\top}\mathbf{V}^{(i)} + \lambda_1 \alpha_i \mathbf{H}^{(i)} + \mathbf{V}^{(i)\top}\mathbf{V}^{(i)} + \lambda_1 \mathbf{H}^{(i)\top}\mathbf{H}^{(i)})). \quad (16)$$

Setting the derivative with respect to $\mathbf{Z}_{:,j}^{(i)}$ to zero, we can get

$$\frac{\partial \mathcal{L}(\mathbf{Z}_{:,j}^{(i)})}{\partial \mathbf{Z}_{:,j}^{(i)}} = \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{Q}^{(i)}\mathbf{Z}_{:,j}^{(i)} + \mathbf{Q}^{(i)\top}\mathbf{Z}_{:,j}^{(i)} - \mathbf{p}^{(i)} = 0,$$

where $\mathbf{Q}^{(i)} = \mathbf{V}^{(i)\top} \mathbf{V}^{(i)} + \lambda_1 \alpha_i^2 \mathbf{I}$ and $\mathbf{p}^{(i)} = 2(\mathbf{V}^{(i)\top} \mathbf{V}^{(i)})_{:,j} + 2\lambda_1 \alpha_i \mathbf{H}_{:,j}^{(i)}$. For each view, we can obtain the following closed-form solution

$$\mathbf{Z}_{:,j}^{(i)} = (\mathbf{Q}^{(i)} + \mathbf{Q}^{(i)\top})^{-1} \mathbf{p}^{(i)} \quad (17)$$

4.2.4. Update α_i

By fixing $\mathbf{U}^{(i)}$, $\mathbf{V}^{(i)}$, $\mathbf{Z}^{(i)}$ and \mathbf{Z}^* and removing other terms, the optimization for α_i can be transformed into solving Equation (18).

$$\min_{\alpha_i} \sum_{i=1}^m \left\| \alpha_i \mathbf{Z}^{(i)} - \mathbf{O}^{(i)} \mathbf{Z}^* \mathbf{O}^{(i)\top} \right\|_F^2 \quad \text{s.t.} \quad \sum_{i=1}^m \alpha_i = 1, \alpha_i \geq 0 \quad (18)$$

Note that $\mathbf{H}^{(i)} = \mathbf{O}^{(i)} \mathbf{Z}^* \mathbf{O}^{(i)\top}$. For each view, we can obtain the following Lagrange function:

$$\mathcal{L}(\alpha_i) = \alpha_i^2 \text{Tr}(\mathbf{Z}^{(i)\top} \mathbf{Z}^{(i)}) - 2\alpha_i \text{Tr}(\mathbf{Z}^{(i)\top} \mathbf{H}^{(i)}) + \text{Tr}(\mathbf{H}^{(i)\top} \mathbf{H}^{(i)}) + \gamma \left(\sum_{i=1}^m \alpha_i - 1 \right), \quad (19)$$

where γ is the Lagrange multipliers. Setting the derivative of $\mathcal{L}(\alpha_i)$ w.r.t. α_i to zero, we can obtain:

$$\alpha_i = \frac{2\text{Tr}(\mathbf{Z}^{(i)\top} \mathbf{H}^{(i)}) - \gamma}{2\text{Tr}(\mathbf{Z}^{(i)\top} \mathbf{Z}^{(i)})} \quad (20)$$

According to the constraint $\sum_{i=1}^m \alpha_i = 1$, we can compute γ and further get each α_i .

4.2.5. Update \mathbf{Z}^*

With $\mathbf{U}^{(i)}$, $\mathbf{V}^{(i)}$, $\mathbf{Z}^{(i)}$ and α_i fixed, we need to minimize the following objective for $\mathbf{Z}^{(i)}$.

$$\min_{\mathbf{Z}^*} \sum_{i=1}^m \lambda_1 \left\| \alpha_i \mathbf{Z}^{(i)} - \mathbf{O}^{(i)} \mathbf{Z}^* \mathbf{O}^{(i)\top} \right\|_F^2 + \lambda_2 \|\mathbf{Z}^*\|_1 \quad (21)$$

We can get the equivalent element-wise equation:

$$\min_{\mathbf{Z}^*} \lambda_1 \left\| \sum_{i=1}^r \alpha_i \mathbf{Z}_{j,p}^{(i)} - r \mathbf{Z}_{k,q}^* \right\|^2 + \lambda_2 |\mathbf{Z}_{k,q}^*| \quad (22)$$

Note that j, p differ in different views. r is the count of views in which samples j and p exist simultaneously. If $O_{j,k}^{(i)} = 1$ and $O_{p,q}^{(i)} = 1$, the elements $\mathbf{Z}_{k,q}^*$ in \mathbf{Z}^* are composed of the weighted sum of the corresponding instance $\mathbf{Z}_{j,p}^{(i)}$ in view i . Specifically, Equation (22) obtains a unique solution:

$$\mathbf{Z}_{k,q}^* = \begin{cases} \frac{2\lambda_1 r \sum_{i=1}^r \alpha_i \mathbf{Z}_{j,p}^{(i)} - \lambda_2}{2\lambda_1 r^2}, & \text{if } \sum_{i=1}^r \alpha_i \mathbf{Z}_{j,p}^{(i)} > \frac{\lambda_2}{2\lambda_1 r} \\ \frac{2\lambda_1 r \sum_{i=1}^r \alpha_i \mathbf{Z}_{j,p}^{(i)} + \lambda_2}{2\lambda_1 r^2}, & \text{if } \sum_{i=1}^r \alpha_i \mathbf{Z}_{j,p}^{(i)} < \frac{-\lambda_2}{2\lambda_1 r} \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

As can be seen in Equation (23), the solution of optimal \mathbf{Z}^* is a weighted combination of self-representation graph over the view which corresponding samples are visible. Moreover, the noise and outliers will be given a very small value to make the \mathbf{Z}^* sparse. Therefore, we can get a robust and complete graph revealing all of the relationships of the samples.

4.3. Convergence and Computational Complexity

We end up in this section by analyzing the convergence analysis and computational complexity of our proposed method.

Convergence analysis: We first analyze the convergence of the proposed method. Algorithm 1 is a convex problem during the updating of each variable. Each sub-problem obtains a global optimum solution and the value of the objective function is non-increasing until converges. Experiment results in the next section demonstrate this in practice.

Computational complexity analysis: With the optimization process outlined in Algorithm 1, the total time complexity consists of five parts referring to the alternate steps. For incomplete multi-view setting, dimensionality and the number of complete samples varies across different views. With notations in the following, the first stage for computing $\mathbf{U}^{(i)}$ needs $\mathcal{O}(d_i n_i k + n_i k^2 + k^3 + d_i k^2)$ for the i -th view. The time cost of $d \times n$ dimensional matrix \mathbf{X} multiplying $n \times k$ dimensional matrix \mathbf{V}^\top is $\mathcal{O}(d_i n_i k)$. Similarly, the time costs of $\mathbf{V}\mathbf{V}^\top$ and $(\mathbf{V}\mathbf{V}^\top)^{-1}$ are $\mathcal{O}(n_i k^2)$ and $\mathcal{O}(k^3)$, respectively. At last, the result of $\mathbf{X}\mathbf{V}^\top$ times $(\mathbf{V}\mathbf{V}^\top)^{-1}$ costs $\mathcal{O}(d_i k^2)$. Therefore the total time cost of updating $\mathbf{U}^{(i)}$ is $\mathcal{O}(d_i n_i k + n_i k^2 + k^3 + d_i k^2)$ for each view. Similarly to updating $\mathbf{U}^{(i)}$, the time cost of updating $\mathbf{V}^{(i)}$ is $\mathcal{O}(q(kn_i^2 + d_i k^2 + k^2 n_i + kd_i n_i))$, where q is the number of iterations. The time cost of updating $\mathbf{Z}^{(i)}$ is $\mathcal{O}(n_i(n_i k^2 + n_i^3))$. The time cost of updating α_i is $\mathcal{O}(mn_i^3)$. At last, solving \mathbf{Z}^* acquires $\mathcal{O}(mn_i^2)$. After all, the time complexity of our algorithm is $\mathcal{O}(mn_i(n_i k^2 + n_i^3 + n_i + n_i k))$.

Algorithm 1: AWGF-IMSC

Input: Multi-view data with m views $\{\mathbf{X}^{(i)}\}_{i=1}^m$, index matrix $\{\mathbf{O}^{(i)}\}_{i=1}^m$, the number of cluster k , hyper-parameters λ_1 and λ_2 .

Initialize: Initialize $\mathbf{V}^{(i)} \in \mathbb{R}^{k \times n_i}$ based on preliminary k -means. If sample i belongs to cluster k , the corresponding element in $\mathbf{V}^{(i)}$ equals to 1 else 0. Initialize $\mathbf{Z}^{(i)}$ with the k -nearest neighbors on visible data points in each view. Initialize α_i with $\frac{1}{m}$, where m is the number of views. Initialize \mathbf{Z}^* with the average of $\mathbf{Z}^{(i)}$.

while not convergence do

Update $\mathbf{U}^{(i)}$ by Equation (11);
 Update $\mathbf{V}^{(i)}$ by solving Equation (14);
 Update $\mathbf{Z}^{(i)}$ by Equation (17);
 Update α_i by Equation (20);
 Update \mathbf{Z}^* by Equation (23);

end

Output: \mathbf{Z}^*

Performing spectral clustering on \mathbf{Z}^* to get the clustering results.

5. Experiment

5.1. Datasets

To demonstrate the effectiveness of our proposed algorithm AWGF_IMSC, we do comparisons with six baseline methods on four benchmark datasets. The statistical information of the datasets is displayed in Table 1. Detailed introductions are as follows:

- BUAA-visnir face database (BUAA) [49]. The dataset BUAA used in this paper contains 1350 instances of 150 categories. Each instance has visible images (VIS) and near infrared images (NIR), which naturally form a two-view dataset. Both VIS and NIR images are 640×480 pixels. Then, they are resized into 10×10 matrix and vectorized into 100-dimensional features.

- Caltech7 [50]. The Caltech7 dataset is a subset of the Caltech101 dataset, containing seven categories (Face, Motorbikes, Dolla-Bill, Garfield, Snoopy, Stop-Sign and Windsorchair) and 1474 instances. The original images of dataset Caltech7 differ in size. We follow the work in [48], selecting two of five given features as the multi-view dataset. The selected two views refer to 512 dimensional GIST features [51] and 928 dimensional local binary patterns(LBP) features [51].
- One-hundred plant species leaves dataset (100Leaves) [52]. The 100Leaves dataset contains 1600 instances from 100 categories. The original images of 100Leaves differ in size, too. Shape descriptor, fine scale margin and texture histogram features constitute three-views to depict samples from different perspectives.
- Mfeat handwritten digit dataset (Mfeat) [53]. This dataset contains 2000 samples. The size of the original images of dataset Mfeat is 891×702 pixels. The public multi-view dataset of it has six views. In our experiments, we select 76-dimensional features of Fourier coefficients of the character shapes and 240-dimensional features of pixel averages.

Table 1. Statistics of the datasets.

	Samples	Views	Clusters	Feature		
BUAA	1350	2	150	100	100	
Caltech7	1474	2	7	512	928	
100Leaves	1600	3	100	64	64	64
Mfeat	2000	2	10	76	240	

As shown in Figure 2, we randomly select six pictures of different categories from four original datasets for display.

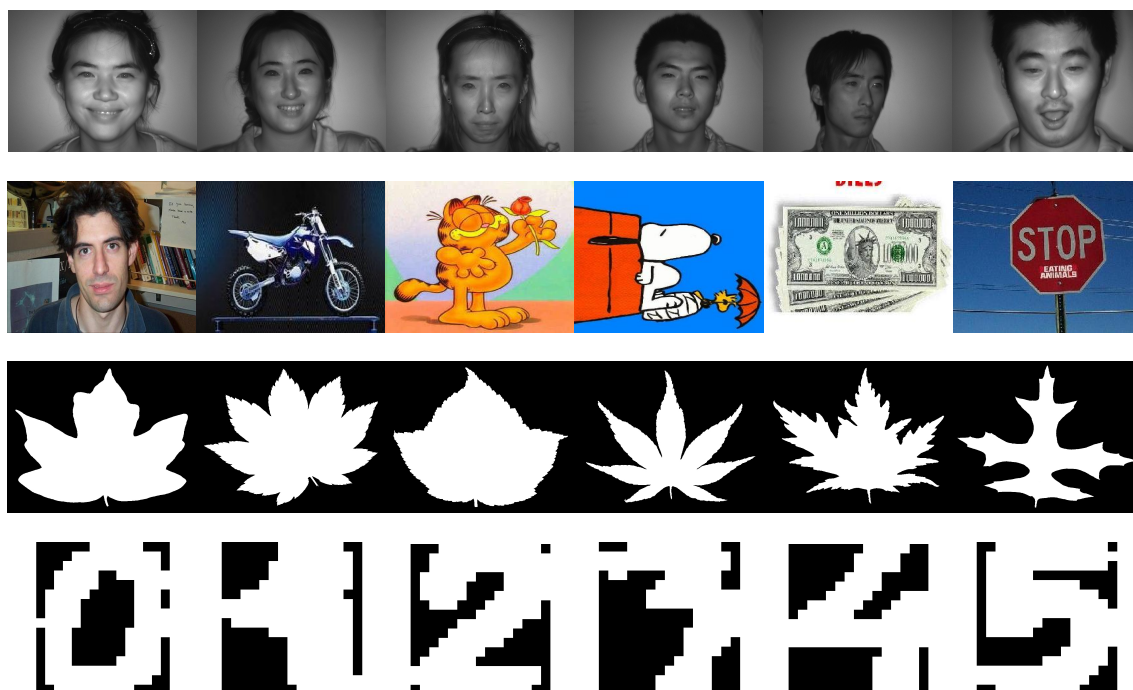


Figure 2. Sample images of the four datasets. Row from top to bottom represents images from BUAA, Caltech7, 100Leaves and Mfeat, respectively.

5.2. Baselines

We conduct extensive experiments, comparing with several state-of-the-art incomplete multi-view clustering methods. Brief introductions are given below.

- Best single view (BSV). BSV first fills the missing samples with the average feature values of its view. The affinity matrices can be constructed by Gaussian kernel. Then, we perform spectral clustering algorithm on the similarity matrix of each view and report the best clustering performance.
- Partial multi-view clustering (PVC) [35]. This method supposes that the instances available in both views should have a common representation. The view-specific instances which are missing in another view should maintain the specific information. Based on the NMF, this method integrates the common and view-specific representations in the latent space to form a unified representation.
- Multiple incomplete view clustering via weighted non-negative matrix factorization with $\ell_{2,1}$ regularization (MIC) [38]. This paper first fills the missing instances with an average value of features and then learns a $\ell_{2,1}$ regularized latent subspace by weighted NMF.
- Incomplete multi-modal visual data grouping (IMG) [36]. IMG proposes to use the latent representation to generate a complete graph, which establishes a connection between missing data from different views.
- Doubly aligned incomplete multi-view clustering (DAIMC) [37]. The proposed method first aligns the samples into a common representation by semi-NMF and then aligns the base matrices with the help of $\ell_{2,1}$ regularized regression model.
- Incomplete multi-view spectral clustering with adaptive graph learning (INMF-AGL) [48] induces a co-regularization term to learn the common representation, which integrates the graph learning and spectral clustering.

For the compared methods, we run their demo with the suggest or default parameters and repeat five times to obtain average results. Note that the PVC and IMG methods can only deal with the two-view scenarios. In our experiments, we combine two different views and report the best result.

Following most existing works, we utilize accuracy (ACC) and normalized mutual information (NMI) to measure the clustering results; higher values representing better clustering performance. Then we give the definition of the two metrics. Denoting true positive (TP), false positive (FP), false negative (FN) and true negative (TN), we can obtain the ACC and NMI definitions as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (24)$$

indicating the percentage of correct predicted results in the total samples.

NMI quantifies the amount of information contained in a random variable about another random variable. For clarity, the expression can be formulated as:

$$NMI(X, Y) = 2 \frac{I(X, Y)}{H(X) + H(Y)} \quad (25)$$

where mutual information $I(X, Y)$ is $\sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$, $p(x, y)$ is the joint probability distribution of X and Y , $p(x)$ is the marginal probability distribution of X . $H(X) = -\sum_i p(x_i) \log p(x_i)$ is the information entropy, regarded as the uncertainty of random variables.

5.3. Experiment Setting

In our experiments, we generate incomplete data from complete multi-view datasets in the way of *One-complete*, which means that we randomly select one of the views to be complete. The rest of the views suffer different incomplete ratio (IR) from 10%, 20%, 30%, 40%, 50%, 60%, 70%. For the two-view datasets BUAA, Caltech7 and Mfeat, we randomly select one view as the complete view. The incomplete case occurs in the rest view with randomly removing 10–70% samples. For more than two-view occasions, one view is chosen randomly and the rest views suffer 10–70% missing. The datasets used in this paper can be found in our Github (<https://github.com/Jeaninezpp/Incomplete-multi-view-datasets>). In our experiments, we perform our proposed method five times as the compared method for fairness. Our code is available at <https://github.com/Jeaninezpp/AWGF-code>.

5.4. Experiment Results and Analysis

Experiment results on different datasets of various compared method are enumerated in Tables 2–5. These four tables present the ACC results of the above algorithms on four benchmark datasets. Each row shows the accuracy of compared methods under a certain incomplete ratio. We highlight the best results in bold. Each column represents the evolution of the accuracy of the corresponding method as the incomplete ratio increase. Under each incomplete ratio, we can get the sequence number by sorting the accuracy from high to low. The sequence number is regarded as the rank of the algorithms under a certain incomplete ratio. Then, we can obtain the average rank by taking the average of the ranks over each algorithm. The average rank illustrates the robustness of the method in terms of incomplete ratio.

Table 2. ACCs and average ranks on Caltech7 under different incomplete ratios. Bold numbers denote the best result.

IR\Method	BSV	MIC	IMG	DAIMC	INMF_AGL	Ours
10%	0.3328	0.4007	0.5189	0.4105	0.5263	0.7605
20%	0.3145	0.3886	0.5027	0.3969	0.5794	0.7104
30%	0.3436	0.3205	0.4837	0.4122	0.5791	0.7293
40%	0.4139	0.3366	0.5080	0.3423	0.6007	0.6981
50%	0.4861	0.3493	0.4290	0.3550	0.5702	0.6750
60%	0.5245	0.3446	0.4943	0.4155	0.5874	0.6954
70%	0.4324	0.3539	0.4837	0.4109	0.5799	0.6942
Average Rank	4.43	5.71	3.29	4.57	2.00	1.00

Table 3. ACCs and average ranks on BUAA under different incomplete ratios. Bold numbers denote the best result.

IR\Method	BSV	MIC	IMG	DAIMC	INMF_AGL	Ours
10%	0.2964	0.0193	0.3424	0.3203	0.5652	0.6446
20%	0.2997	0.0193	0.3424	0.2775	0.5637	0.6395
30%	0.3006	0.0193	0.3424	0.2341	0.5566	0.6284
40%	0.2997	0.0193	0.3424	0.2336	0.5505	0.6258
50%	0.2965	0.0193	0.3424	0.2373	0.4730	0.6223
60%	0.2975	0.0193	0.3424	0.2413	0.5293	0.6217
70%	0.2979	0.0193	0.3424	0.2596	0.5410	0.6110
Average Rank	4.14	6.00	3.00	4.86	2.00	1.00

Table 4. ACCs and average ranks on 100Leaves under different incomplete ratios. Bold numbers denote the best result.

IR\Method	BSV	MIC	IMG	DAIMC	INMF_AGL	Ours
10%	0.1069	0.6208	0.5661	0.6628	0.8223	0.8433
20%	0.1064	0.5768	0.5143	0.5750	0.7835	0.8104
30%	0.1060	0.5234	0.4607	0.4740	0.7529	0.7776
40%	0.1056	0.4744	0.4302	0.4370	0.7026	0.7291
50%	0.1075	0.4665	0.4001	0.3068	0.6551	0.7028
60%	0.1065	0.4449	0.3668	0.3405	0.6316	0.6478
70%	0.1057	0.4261	0.3740	0.3875	0.5893	0.6194
Average Rank	6.00	3.14	4.71	4.14	2.00	1.00

Table 5. ACCs and average ranks on Mfeat under different incomplete ratios. Bold numbers denote the best result.

IR\Method	BSV	MIC	IMG	DAIMC	INMF_AGL	Ours
10%	0.1520	0.6477	0.5442	0.8670	0.8650	0.8100
20%	0.1484	0.5721	0.5004	0.7151	0.8415	0.7866
30%	0.1471	0.5220	0.5133	0.5737	0.8177	0.7995
40%	0.1497	0.4539	0.4554	0.5042	0.8148	0.7864
50%	0.1458	0.3640	0.4037	0.5423	0.7989	0.7680
60%	0.1524	0.3583	0.3424	0.5819	0.7214	0.7515
70%	0.1476	0.3447	0.3424	0.6446	0.7027	0.7256
Average Rank	6.00	4.29	4.71	2.71	1.43	1.86

Furthermore, we depict the NMI metric in Figure 3 with line charts. Based on these results, we have the following observations:

- Compared with the proposed method, the BSV method yields worse clustering performance. This is mainly because directly filling the missing instances with the average features will lead them to be clustered into the same group. The weighted NMF methods DAIMC and MIC perform better than BSV at a low missing rate since the NMF-based methods learn a shared representation to exploit the complementary information across views. Besides, the weighted manner reduces the negative impact of the missing instances. However, with the increasing incomplete ratio, these two methods suffer a sharp decline, especially apparent in Mfeat dataset. Methods like IMG and INMF_AGL involving the graph construction perform better than them. Our proposed method integrates the advantages of NMF-based and graph-based methods, adaptively fusing the graph learned from each embedding space. Therefore, our AWGF_IMSC method reaches the best clustering performance in most cases.
- Comparing with the INMF_AGL [48], our proposed method consistently further improves the clustering performance and achieves better results among the benchmark datasets. In addition to the cases on Mfeat, the INMF_AGL method performs higher accuracy than AWGF_IMSC under the 20%-50% missing rate. However, our performance exceeds it under the incomplete ratio in 60% and 70%. Although both of INMF_AGL and our AWGF_IMSC adopt subspace clustering to build graph structure in each view, the clustering results demonstrate the effectiveness of graph fusion instead of indicator fusion in INMF_AGL.
- AWGF-IMSC shows clear advantages over other compared baselines under various incomplete ratios, with three best and one second-best results out of the total four datasets. For example, on the BUAA dataset (Table 3), our method transcends the second best method by 7.94%, 7.58%, 7.18%, 7.53%, 14.93%, 9.24% and 7%, respectively. More significant improvements can be seen on dataset Caltech7. In Table 2, the ACCs of the proposed method are 23.42%, 13.10%, 15.02%, 9.74%, 10.48%, 10.8% and 11.43% higher than the second best INMF_AGL method. These significant results verify the effectiveness of the proposed adaptive weighted graph-based fusion learning for incomplete multi-view clustering. Our method achieves the best average rank in Caltech7, BUAA and 100Leaves. In Mfeat, the average rank of the proposed method is second best, which is only 0.43 more than the best, but 0.85 less than the third.
- As shown in Figure 3, we can also observe that the proposed algorithm outperforms other methods on all of the datasets under various incomplete ratio. Besides, our method appears a relatively stable trend as the missing rate increases. Moreover, the abnormal phenomenon of BSV in dataset Caltech7 (Figure 3a) maybe because the preserved complete view has an excellent structure when generating large missing datasets. Therefore, the results of the BSV will be outstanding. Other methods are affected by the negative impact of missing samples and thus produce a lower effect than BSV, while our method is still superior to all compared methods by a more significant proportion, further illustrating the effectiveness and superiority of the proposed method.

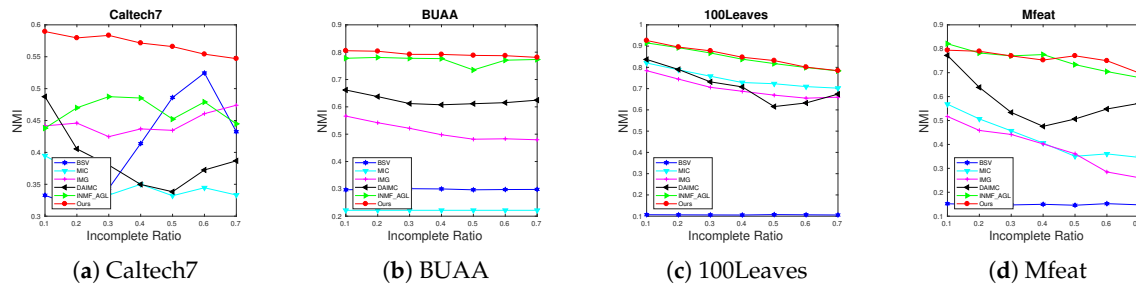


Figure 3. NMI results with different incomplete ratios on four incomplete datasets.

5.5. Analysis of the Parameter Sensitivity

In this section, we analyze the impact of the hyper-meters λ_1 and λ_2 in AWGF-IMSC on clustering performance. The parameters are chosen ranging from $[10^{-3}, 10^{-2}, \dots, 10^3]$ by grid search. Figures 4 and 5 plot the NMI results by varying λ_1 and λ_2 in a large range on BUAA, Caltech7, 100Leaves and Mfeat.

We have the following observations from Figures 4 and 5: (i) All the two hyper-parameters are effective in improving the clustering performance; (ii) AWGF-IMSC is practically stable against these parameters that it achieves competitive performance in a wide range of parameter settings in a low missing rate; (iii) With the incomplete ratio increasing, the combinations of a relatively smaller λ_1 and a more prominent λ_2 tend to achieve better performance. The reason is that the λ_2 controls the impact of the sparse regularizer of the complete graph. Imposing sparseness requirements on the graph within a specific range will filter out the noise in the graph and the inconsistency between the views. (iv) For different datasets, we can conclude that most of the datasets are relatively stable when $\lambda_1 \in \{1e-3, 1e-2, 1e-1\}$, $\lambda_2 \in \{10, 100, 1000\}$.

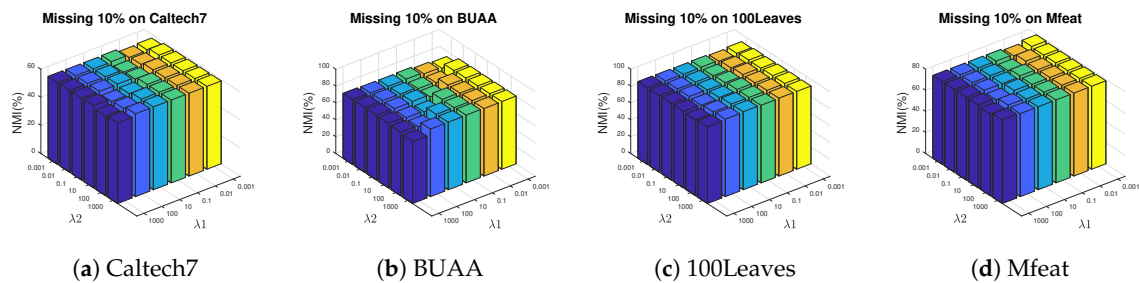


Figure 4. Parameter study on four incomplete datasets. Pictures depict the diversification of NMI under different parameter combinations in the condition of missing 10% instances.

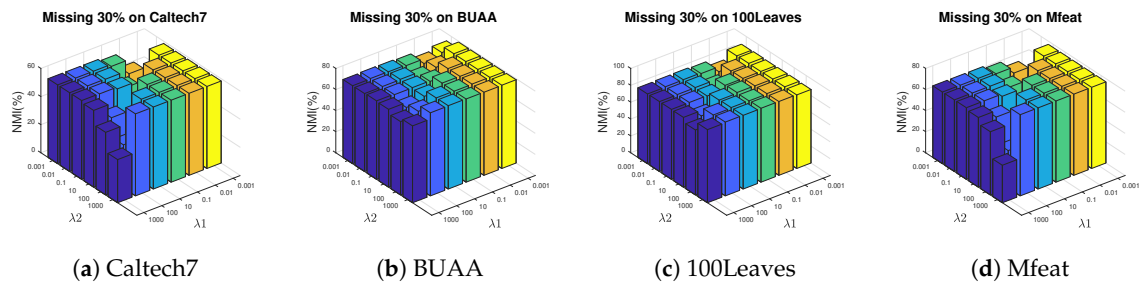


Figure 5. The influence of parameters λ_1 and λ_2 on NMI when 30% instances are missing on four datasets.

5.6. Convergence Analysis

Our algorithm is theoretically guaranteed to converge to a local minimum, as illustrated in the optimization part. For the convergence analysis, we conduct experiments on the four datasets with all incomplete ratios and all suggest parameter scope. We randomly select from each dataset and draw the evolution of the objective value, as shown in Figure 6. In the above experiments, we observe that our algorithm's objective values monotonically decrease at each iteration and usually converge in less than 20 iterations. These results verify our proposed algorithm's convergence.

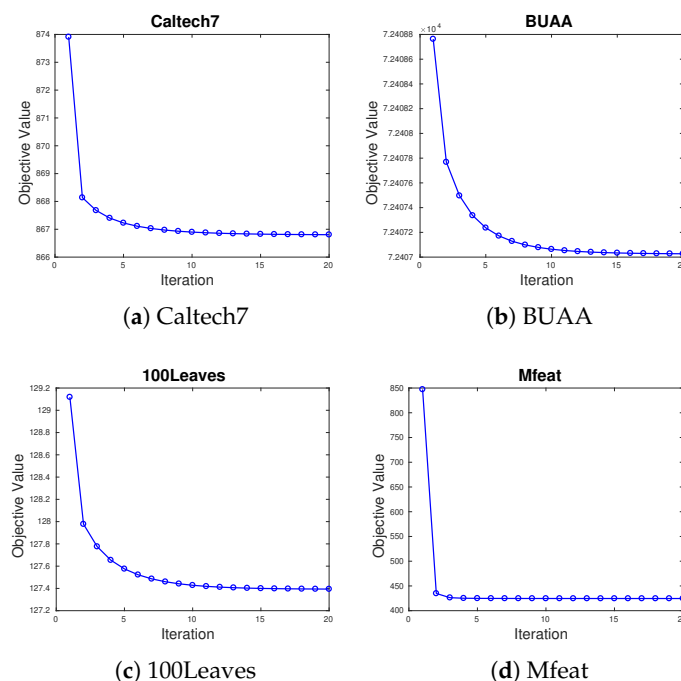


Figure 6. The convergence curves of the objective values on four datasets.

6. Conclusions

This article proposes a novel incomplete multi-view clustering method to fuse the local-structure contained graph with adaptive view-importance learning. We incorporate representation learning and incomplete graph fusion into a unified framework, whereas two processes can negotiate with each other, serving for graph learning tasks. A sparse regularization is imposed on the complete graph to do it more robustly to the view-inconsistency. Moreover, the importance of different views is automatically learned, further guiding the construction of the complete graph. We conduct experiments to illustrate the effectiveness and superiority of the proposed method. Some recent works utilize deep neural networks, such as GAN (generative adversarial network), to generate missing features to solve incomplete multi-view clustering problems. The utilization of neural networks to handle multi-view clustering and incomplete multi-view clustering will be advanced considerations in the future.

Author Contributions: All authors contributed to the manuscript preparation of the paper. Conceptualization, P.Z. and S.W.; methodology, P.Z.; validation, P.Z. and J.H.; formal analysis, P.Z. and Z.C. (Zhen Cheng); investigation, P.Z. and X.G.; data curation, P.Z.; writing—original draft preparation, P.Z.; writing—review and editing, S.W., J.H., Z.C. (Zhen Cheng) and P.C. (Zhiping Cai); visualization, P.Z. and X.G.; supervision, E.Z., P.C. (Zhiping Cai) and S.W.; funding acquisition, E.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China 2018YFB1003203 and the National Natural Science Foundation of China (Grant NO.61672528, NO. 61773392, NO. 62072465).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AWGF-IMSC	Adaptive weighted graph fusion incomplete multi-view subspace clustering
MVC	Multi-view clustering
NMF	Non-negative matrix factorization
BUAA	BUAA-visnir face database
100Leaves	One-hundred plant species leaves dataset
Mfeat	Mfeat handwritten digit dataset
VIS	Visual image
NIR	Near infrared image
LBP	Local binary patterns
IR	Incomplete ratio
ACC	Accuracy
NMI	Normalized mutual information
TP, FP, FN, TN	True positive, false positive, false negative, true negative
BSV	Best single view
PVC	Partial multi-view clustering
MIC	Multiple incomplete view clustering via weighted non-negative matrix factorization with $\ell_{2,1}$ regularization
IMG	Incomplete multi-modal visual data grouping
DAIMC	Doubly aligned incomplete multi-view clustering
INMF-AGL	Incomplete multiview spectral clustering with adaptive graph learning

References

1. Zhao, Q.; Zhang, Y.; Qin, Q.; Luo, B. Quantized Residual Preference Based Linkage Clustering for Model Selection and Inlier Segmentation in Geometric Multi-Model Fitting. *Sensors* **2020**, *20*, 3806.
2. Biabani, M.; Fotouhi, H.; Yazdani, N. An Energy-Efficient Evolutionary Clustering Technique for Disaster Management in IoT Networks. *Sensors* **2020**, *20*, 2647.
3. Deng, T.; Ye, D.; Ma, R.; Fujita, H.; Xiong, L. Low-rank local tangent space embedding for subspace clustering. *Inf. Sci.* **2020**, *508*, 1–21.
4. Peng, X.; Feng, J.; Zhou, J.T.; Lei, Y.; Yan, S. Deep subspace clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**. doi:10.1109/TNNLS.2020.2968848.
5. Chen, J.; Zhao, Z.; Ye, J.; Liu, H. Nonlinear Adaptive Distance Metric Learning for Clustering. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07), San Jose, CA, USA, 12–15 August 2007; Association for Computing Machinery: New York, NY, USA, 2007; pp. 123–132. doi:10.1145/1281192.1281209.
6. Gönen, M.; Alpaydin, E. Localized Multiple Kernel Learning. In Proceedings of the 25th International Conference on Machine Learning (ICML'08), Helsinki, Finland, 5–9 July 2008; Association for Computing Machinery: New York, NY, USA, 2008; pp. 352–359. doi:10.1145/1390156.1390201.
7. Chaudhuri, K.; Kakade, S.M.; Livescu, K.; Sridharan, K. Multi-View Clustering via Canonical Correlation Analysis. In Proceedings of the 26th Annual International Conference on Machine Learning (ICML'09), Montreal, QC, Canada, 14–18 June 2009; Association for Computing Machinery: New York, NY, USA, 2009; pp. 129–136. doi:10.1145/1553374.1553391.
8. Yu, S.; Tranchevent, L.; Liu, X.; Glanzel, W.; Suykens, J.A.; De Moor, B.; Moreau, Y. Optimized data fusion for kernel k-means clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1031–1039. doi:10.1007/978-3-642-19406-1_4.
9. Huang, H.C.; Chuang, Y.Y.; Chen, C.S. Multiple kernel fuzzy clustering. *IEEE Trans. Fuzzy Syst.* **2012**, *20*, 120–134.
10. Liu, X.; Wang, L.; Yin, J.; Zhu, E.; Zhang, J. An efficient approach to integrating radius information into multiple kernel learning. *IEEE Trans. Cybern.* **2013**, *43*, 557–569.

11. Gönen, M.; Margolin, A.A. Localized Data Fusion for Kernel k-Means Clustering with Application to Cancer Biology. In *Advances in Neural Information Processing Systems 27*; Curran Associates, Inc.: New York, NY, USA, 2014; pp. 1305–1313.
12. Liu, X.; Dou, Y.; Yin, J.; Wang, L.; Zhu, E. Multiple Kernel k-Means Clustering with Matrix-Induced Regularization. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 1888–1894.
13. Wang, S.; Zhu, E.; Hu, J.; Li, M.; Zhao, K.; Hu, N.; Liu, X. Efficient multiple kernel k-means clustering with late fusion. *IEEE Access* **2019**, *7*, 61109–61120.
14. Chowdhary, C.L.; Mittal, M.; Pattanaik, P.; Marszalek, Z. An Efficient Segmentation and Classification System in Medical Images Using Intuitionist Possibilistic Fuzzy C-Mean Clustering and Fuzzy SVM Algorithm. *Sensors* **2020**, *20*, 3903.
15. Du, L.; Zhou, P.; Shi, L.; Wang, H.; Fan, M.; Wang, W.; Shen, Y.D. Robust Multiple Kernel K-Means Using $\ell_{2,1}$ Norm. In Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015; pp. 3476–3482.
16. Zhou, S.; Liu, X.; Li, M.; Zhu, E.; Liu, L.; Zhang, C.; Yin, J. Multiple kernel clustering with neighbor-kernel subspace segmentation. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 1351–1362.
17. Wang, S.; Liu, X.; Zhu, E.; Tang, C.; Liu, J.; Hu, J.; Xia, J.; Yin, J. Multi-view Clustering via Late Fusion Alignment Maximization. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 3778–3784. doi:10.24963/ijcai.2019/524.
18. Kumar, A.; III, H.D. A Co-Training Approach for Multi-View Spectral Clustering. In Proceedings of the 28th International Conference on Machine Learning (ICML'11), Bellevue, WA, USA, 28 June–2 July 2011; Omnipress: Madison, WI, USA, 2011; pp. 393–400.
19. Kumar, A.; Rai, P.; Daumé, H. Co-Regularized Multi-View Spectral Clustering. In Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS'11), Guangzhou, China, 14–18 November 2017; Curran Associates Inc.: Red Hook, NY, USA, 2011; pp. 1413–1421.
20. Zhang, Z.; Liu, L.; Shen, F.; Shen, H.T.; Shao, L. Binary multi-view clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 1774–1782.
21. Jiang, G.; Wang, H.; Peng, J.; Chen, D.; Fu, X. Graph-based Multi-view Binary Learning for Image Clustering. *arXiv* **2019**, arXiv:1912.05159.
22. Zhang, C.; Fu, H.; Hu, Q.; Cao, X.; Xie, Y.; Tao, D.; Xu, D. Generalized latent multi-view subspace clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 86–99.
23. Zhang, C.; Hu, Q.; Fu, H.; Zhu, P.; Cao, X. Latent Multi-view Subspace Clustering. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE Computer Society: Los Alamitos, CA, USA, 2017; pp. 4333–4341. doi:10.1109/CVPR.2017.461.
24. Brbić, M.; Kopriva, I. Multi-view low-rank sparse subspace clustering. *Pattern Recognit.* **2018**, *73*, 247–258.
25. Kang, Z.; Zhao, X.; Peng, C.; Zhu, H.; Zhou, J.T.; Peng, X.; Chen, W.; Xu, Z. Partition level multiview subspace clustering. *Neural Netw.* **2020**, *122*, 279–288.
26. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust Recovery of Subspace Structures by Low-Rank Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 171–184.
27. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791.
28. Liu, J.; Wang, C.; Gao, J.; Han, J., Multi-View Clustering via Joint Nonnegative Matrix Factorization. In Proceedings of the 2013 SIAM International Conference on Data Mining, Austin, TX, USA, 2–4 May 2013; pp. 252–260. doi:10.1137/1.9781611972832.28.
29. Zong, L.; Zhang, X.; Zhao, L.; Yu, H.; Zhao, Q. Multi-view clustering via multi-manifold regularized non-negative matrix factorization. *Neural Networks* **2017**, *88*, 74–89.
30. Yang, Z.; Liang, N.; Yan, W.; Li, Z.; Xie, S. Uniform Distribution Non-Negative Matrix Factorization for Multiview Clustering. *IEEE Trans. Cybern.* **2020**.
31. Yin, M.; Huang, W.; Gao, J. Shared Generative Latent Representation Learning for Multi-View Clustering. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 6688–6695.
32. Zhao, L.; Chen, Z.; Yang, Y.; Wang, Z.J.; Leung, V.C. Incomplete multi-view clustering via deep semantic mapping. *Neurocomputing* **2018**, *275*, 1053–1062.

33. Yu, X.; Li, H.; Zhang, Z.; Gan, C. The Optimally Designed Variational Autoencoder Networks for Clustering and Recovery of Incomplete Multimedia Data. *Sensors* **2019**, *19*, 809.
34. Liu, X.; Li, M.; Tang, C.; Xia, J.; Xiong, J.; Liu, L.; Kloft, M.; Zhu, E. Efficient and effective regularized incomplete multi-view clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. doi:10.1109/TPAMI.2020.2974828.
35. Li, S.Y.; Jiang, Y.; Zhou, Z.H. Partial Multi-View Clustering. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014; pp. 1968–1974.
36. Zhao, H.; Liu, H.; Fu, Y. Incomplete Multi-Modal Visual Data Grouping. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 2392–2398.
37. Hu, M.; Chen, S. Doubly Aligned Incomplete Multi-View Clustering. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 2262–2268.
38. Shao, W.; He, L.; Yu, P.S. Multiple Incomplete Views Clustering via Weighted Nonnegative Matrix Factorization with $\ell_{2,1}$ Regularization. In Proceedings of the 2015th European Conference on Machine Learning and Knowledge Discovery in Databases, Porto, Portugal, 7–11 September 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 318–334. doi:10.1007/978-3-319-23528-8_20.
39. Gao, H.; Peng, Y.; Jian, S. Incomplete multi-view clustering. In *International Conference on Intelligent Information Processing*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 245–255. doi:10.1007/978-3-319-48390-0_25.
40. Wang, J.; Lai, S.; Li, M. Improved image fusion method based on NSCT and accelerated NMF. *Sensors* **2012**, *12*, 5872–5887.
41. Zhou, Q.; Feng, Z.; Benetos, E. Adaptive noise reduction for sound event detection using subband-weighted NMF. *Sensors* **2019**, *19*, 3206.
42. Ding, C.H.Q.; Li, T.; Jordan, M.I. Convex and Semi-Nonnegative Matrix Factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 45–55. doi:10.1109/TPAMI.2008.277.
43. He, M.; Yang, Y.; Wang, H. Learning latent features for multi-view clustering based on NMF. In *International Joint Conference on Rough Sets*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 459–469.
44. Hu, M.; Chen, S. One-pass incomplete multi-view clustering. In Proceedings of the AAAI Conference on Artificial Intelligence 2019, Honolulu, HI, USA, 27–28 January 2019; Volume 33, pp. 3838–3845.
45. Parsons, L.; Haque, E.; Liu, H. Subspace clustering for high dimensional data: a review. *Acm Sigkdd Explor. Newsl.* **2004**, *6*, 90–105.
46. Hoppenstedt, B.; Reichert, M.; Kammerer, K.; Probst, T.; Schlee, W.; Spiliopoulou, M.; Pryss, R. Dimensionality Reduction and Subspace Clustering in Mixed Reality for Condition Monitoring of High-Dimensional Production Data. *Sensors* **2019**, *19*, 3903.
47. Elhamifar, E.; Vidal, R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2765–2781.
48. Wen, J.; Xu, Y.; Liu, H. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Trans. Cybern.* **2018**, 1418–1429. doi:10.1109/TCYB.2018.2884715.
49. Huang, D.; Sun, J.; Wang, Y. *The Buaa-Visnir Face Database Instructions*; Technical Report IRIP-TR-12-FR-001; School Computer Science and Engineering, Beihang University: Beijing, China, 2012.
50. Fei-Fei, L.; Fergus, R.; Perona, P. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 27 June–2 July 2004.
51. Oliva, A.; Torralba, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vis.* **2001**, *42*, 145–175.
52. Mallah, C.; Cope, J.; Orwell, J. Plant leaf classification using probabilistic integration of shape, texture and margin features. *Signal Process. Pattern Recognit. Appl.* **2013**, *5*, 45–54.
53. van Breukelen, M.; Duin, R.P.; Tax, D.M.; Den Hartog, J. Handwritten digit recognition by combined classifiers. *Kybernetika* **1998**, *34*, 381–386.

