

Orthogonal Spatial-Aware Multi-View Anchor Graph Clustering for Incomplete Remote Sensing Data

Yongshan Zhang^{1,*}, Xiaohuan Lin², Lefei Zhang³, Zhihua Cai²

¹School of Software Engineering, Sun Yat-sen University, Zhuhai, China

²School of Computer Science, China University of Geosciences, Wuhan, China

³School of Computer Science, Wuhan University, Wuhan, China

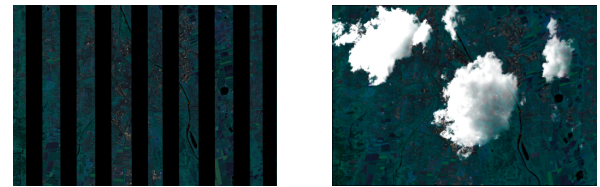
zhangyush95@mail.sysu.edu.cn, {lxhuan, zhcai}@cug.edu.cn, zhangleifei@whu.edu.cn

Abstract

Multi-view clustering for remote sensing data has received increasing attention by leveraging multiple representations to enhance Earth observation. Existing methods are primarily developed under the assumption that each pixel is fully observed across all views. No prior work has investigated the more practical yet challenging scenario where some views suffer from partially missing data. To bridge this gap, we propose the first framework for incomplete multi-view remote sensing data clustering, termed orthogonal spatial-aware multi-view anchor graph clustering (OSMAGC). Specifically, spatial-aware anchors and multi-scale anchor graphs are initially constructed by exploiting the superpixel-based texture characteristics of each view. Based on these, multi-scale anchor graph learning is performed through view weighting and matrix factorization on incomplete data. Structure-aligned consensus feature learning is achieved by aligning the multi-scale graph structures within a shared latent space. To ensure spatial continuity and smoothness, orthogonal spatial-aware regularization is imposed in both horizontal and vertical directions. The latter three modules are jointly optimized through a well-designed optimization algorithm in a mutually reinforcing manner. Extensive experiments on four benchmark datasets validate the effectiveness and efficiency of our proposed method over the state-of-the-art competitors. Code repository: <https://github.com/ZhangYongshan/OSMAGC>.

1. Introduction

Benefiting from advanced techniques, remote sensing data are often described by multiple representations. These multi-view observations provide complementary information from diverse perspectives, facilitating critical Earth observation



(a) Sensor failure.

(b) Cloud obscuration.

Figure 1. Examples and causes of incomplete remote sensing data.

tasks such as urban planning, precision agriculture and military surveillance [1, 7, 41]. Multi-view learning plays a vital role in effectively integrating multiple representations to extract rich semantic cues for improved scene understanding [31, 35]. Given that pixel-level annotation is labor-intensive and time-consuming, especially for large-scale and complex urban scenes, multi-view clustering for remote sensing data has gained increasing attention by exploiting cross-view consistency and complementarity to partition pixels into distinct clusters without requiring ground-truth labels [25, 30].

There are a variety of effective multi-view clustering methods specifically designed for remote sensing data [2, 39, 40]. Based on the nature of view information, existing methods can be categorized into multi-feature-based and multi-source-based methods [8]. Multi-feature-based methods improve clustering performance by leveraging diverse descriptors, such as texture and contour features [7, 30, 38]. However, they exploit multi-view features derived from single-sensor data, providing limited information compared to the heterogeneous data obtained from multiple sensors. In contrast, multi-source-based methods integrate multi-view data from different sensors, such as hyperspectral (HS), multispectral (MS), digital surface model (DSM) and synthetic aperture radar (SAR) images, providing richer and more complementary information for clustering [8, 39]. Although both types of methods can effectively address the clustering problem of multi-view remote sensing data, they encounter

*Corresponding author.

difficulties when dealing with incomplete data [20], where certain views contain missing information caused by sensor failures or cloud occlusion, as illustrated in Figure 1.

To address this issue, incomplete multi-view clustering provides an effective solution by utilizing available data from all views to recover missing content and perform clustering [10, 16, 25]. A variety of incomplete multi-view clustering methods have been proposed by employing various strategies for data completion, such as imputation [16, 29], graph reconstruction [14, 28] and matrix factorization [11, 33]. However, existing methods are not well-suited for remote sensing data, as they fail to effectively exploit the inherent spatial-spectral structures. To enable incomplete multi-view remote sensing data clustering, three critical problems should be considered. (1) Discrepancies in view-specific spectral information provide rich and diverse cues for describing remote sensing data from different perspectives [37]. This raises the challenge of *how to effectively leverage complementary information across incomplete views*. (2) Consensus in land-cover feature representations and distributions across views embodies essential consistency for remote sensing data [3]. This exhibits the challenge of *how to properly capture consistent information under incomplete scenarios*. (3) Missing data within individual views may disrupt spatial coherence and introduce inconsistencies across views [6]. This highlights the challenge of *how to ensure spatial and spectral continuity for more accurate clustering*.

Motivated by these observations, in this paper, we propose an orthogonal spatial-aware multi-view anchor graph clustering (OSMAGC) method for incomplete remote sensing data. As illustrated in Figure 2, the proposed framework consists of four modules. Specifically, spatial-aware anchors and multi-scale anchor graphs are constructed based on the superpixel-based texture information of each view. Then, complete multi-scale anchor graphs are learned from the incomplete data via view weighting and matrix factorization. The structure-aligned consensus features are obtained by aligning multi-scale graph structures into a shared latent space. To enforce spatial consistency, an orthogonal spatial-aware regularizer is introduced along both horizontal and vertical directions. An alternating optimization algorithm is designed to jointly optimize the last three modules in a mutual reinforcement manner. Experiments on four benchmark datasets verify the effectiveness and efficiency of our proposed method under different missing ratios. The main contributions of our work are summarized as follows.

- We propose the first incomplete multi-view clustering method designed for remote sensing data. It is carefully crafted to exploit the complementary and consistent information within their spatial-spectral structures.
- We learn complete multi-scale anchor graphs from incomplete data through view-weighted matrix factorization. This process is guided by high-quality anchors and anchor

graphs initialized using view-specific spatial textures.

- We obtain consensus features through the structure alignment of multi-scale graphs and impose a regularizer to enforce their spatial smoothness. This enables effective aggregation of discriminative information.
- An alternating optimization algorithm is designed to solve the proposed formulation. Experiments demonstrate the superiority of our proposed method.

2. Related Work

Multi-View Clustering for Remote Sensing Data. Due to the abundance of data representations and the lack of labels, numerous multi-view clustering methods have been proposed for remote sensing data to achieve pixel partitioning [2, 21, 36]. These methods can be broadly categorized into two types: multi-feature-based and multi-sensor-based methods. Multi-feature-based methods perform clustering on multiple features extracted from single-sensor remote sensing data. Representative methods include contrastive multiview subspace clustering based on graph convolutional networks (CMSCGC) [7], enhanced multi-view contrastive clustering (EMVCC) [18] and tensorial polarimetric-spectral multiview subspace clustering (TPS-MSC) [5]. In contrast, multi-sensor-based methods integrate remote sensing data from different sensors for clustering. Notable examples include dual graph learning affinity propagation (DGLAP) [39], fast projected fuzzy clustering with anchor guidance (FPFC) [40] and structure-adaptive multi-view graph clustering (SAMVGC) [8]. The aforementioned methods achieve promising results on multi-view remote sensing data. However, they suffer from significant performance degradation when some views contain partially missing data.

Incomplete Multi-View Clustering. Under incomplete multi-view data, numerous methods have been proposed using various effective strategies for data completion and clustering, including imputation, graph reconstruction and matrix factorization [15, 22, 26]. Imputation-based methods fill in missing data to convert the task into a complete clustering problem. Typical methods include incomplete multi-view clustering with reconstructed views (IMCRV) [32] and efficient and effective incomplete multi-view clustering (EE-IMVC) [16]. Graph reconstruction-based methods rebuild a similarity matrix using available data for graph partitioning and clustering. Promising methods include anchor-based incomplete multi-view clustering with cross-view reconstruction (AIMC-CVR) [14] and incomplete multi-view clustering based on consensus graph learning (HCLS-CGL) [27]. Matrix factorization-based methods project available data into a latent space to learn unified representations for clustering. Notable methods include localized sparse incomplete multi-view clustering (LSIMVC) [10] and graph regularized and feature aware matrix factorization (GreatF) [28]. The above-mentioned methods show superior results on general

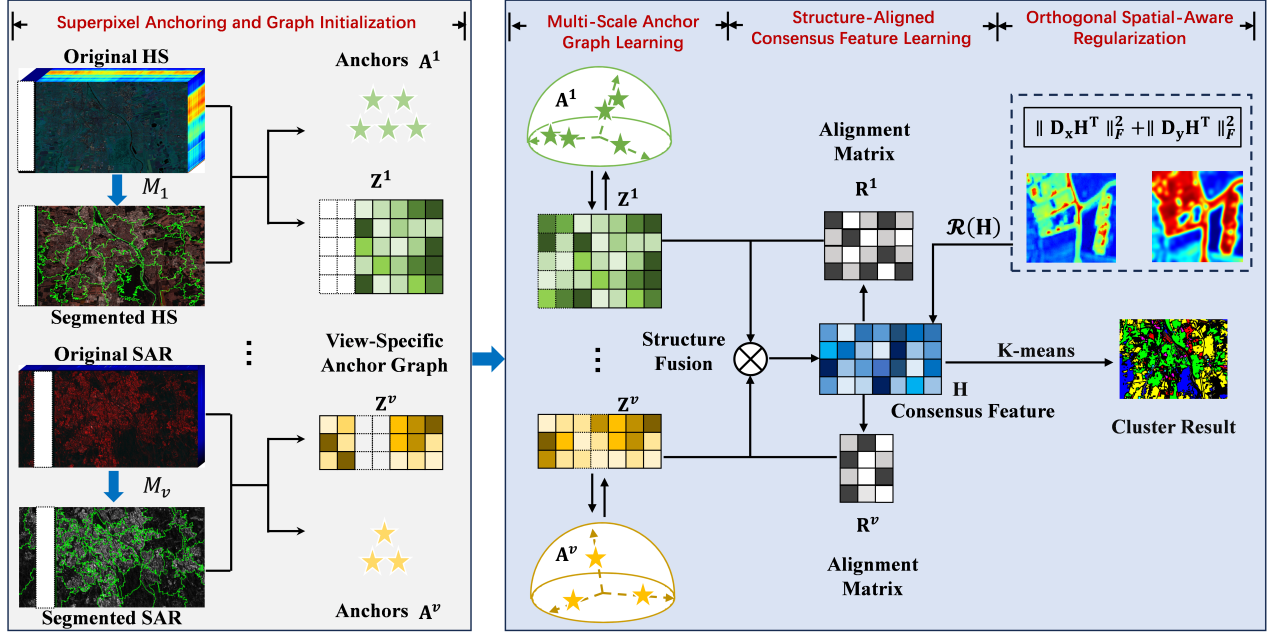


Figure 2. Illustration of our proposed OSMAGC.

incomplete data. However, they perform poorly on incomplete remote sensing data because they fail to capture their spatial-spectral structures.

3. Method

3.1. Problem Definition

Given a multi-view remote sensing dataset $\{\mathbf{X}^v\}_{v=1}^V$ across V views, the v -th view data matrix $\mathbf{X}^v = \{\mathbf{x}_1^v, \dots, \mathbf{x}_N^v\} \in \mathbb{R}^{B_v \times N}$ comprises $N = H \times W$ pixels, where H and W represent the spatial height and width, respectively, and B_v denotes the number of spectral channels. In the context of incomplete data scenarios, a subset of pixels may be missing in each view, and the missing patterns vary across different views. Let $\mathbf{w}^v \in \mathbb{R}^{N_v}$ denote the index vector that indicates the positions of N_v available pixels in the v -th view. Based on this, a binary index matrix $\mathbf{Q}^v \in \{0, 1\}^{N \times N_v}$ is constructed, where each element is defined as $q_{ij}^v = 1$ if $w_j^v = i$, and $q_{ij}^v = 0$ otherwise. Thus, $\mathbf{X}^v \mathbf{Q}^v \in \mathbb{R}^{B_v \times N_v}$ contains all available pixels in the v -th view. For pixels missing in any view, zero-padding is applied to preserve structural consistency. Our goal is to develop an effective method specifically designed to address the challenges of clustering incomplete multi-view remote sensing data.

3.2. Formulation

Based on the initialized spatial-aware anchors and multi-scale anchor graphs, our proposed OSMAGC is formulated as a unified framework that integrates multi-scale anchor graph learning, structure-aligned consensus feature learn-

ing and orthogonal spatial-aware regularization. Figure 2 illustrates the overall framework of OSMAGC.

Superpixel Anchoring and Graph Initialization (SAGI). Spatial-spectral structures are inherent to remote sensing data, where spatial textures have emerged as a crucial factor in anchor graph initialization and learning [40]. However, discrepancies in view-specific spectral information lead to variations in spatial textures across views. To fully exploit the inherent spatial-spectral structures through graph-based representations, anchor graphs should be constructed with adaptive sizes and topologies, each tailored to the spectral and spatial characteristics of the corresponding view. To this end, the entropy rate superpixel (ERS) segmentation technique [12] is individually applied to the first principal component of the v -th view image, taking advantage of its balance between effectiveness and efficiency to extract M_v superpixels based on spatial textures. A set of M_v spatial-aware anchors $\mathbf{A}^v = \{\mathbf{a}_1^v, \dots, \mathbf{a}_{M_v}^v\} \in \mathbb{R}^{B_v \times M_v}$ is then obtained by averaging all pixels within each superpixel to uncover the spatial distribution of pixels. Based on the above preprocessing, an anchor graph $\mathbf{Z}^v \in \mathbb{R}^{M_v \times N}$ is constructed to capture the relationships between M_v anchors and N pixels. Specifically, each element in $\mathbf{Z}^v$ is computed by

$$z_{ij}^v = \begin{cases} \frac{d_{k+1,j}^v - d_{ij}^v}{kd_{k+1,j}^v - \sum_{i=1}^k d_{ij}^v}, & \forall i \in \Phi_j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where Φ_j denotes the set of the k nearest anchors associated with pixel $\mathbf{x}_j^v$, and $d_{ij}^v = \|\mathbf{x}_j^v - \mathbf{a}_i^v\|^2$ indicates the Euclidean distance between $\mathbf{x}_j^v$ and anchor $\mathbf{a}_i^v$. In this way, $\mathbf{Z}^v \in$

$\mathbb{R}^{M_v \times N}$ is constructed in a column-wise manner, where each column characterizes the relationships between a pixel and its neighboring anchors. It is initialized with spatial and spectral information to capture inter-view heterogeneity, facilitating subsequent high-quality learning.

Multi-Scale Anchor Graph Learning (MSAGL). Compared with traditional graph-based methods, anchor graph strategies are better suited for incomplete multi-view clustering, as they consider only the relationships between samples and anchors, thereby enhancing computational efficiency and scalability [19]. Most existing methods typically construct fixed-size anchor graphs and treat all views equally, without considering view-specific characteristics and contributions [24]. To better account for view heterogeneity, anchor graph learning is conducted based on the previously initialized multi-scale anchors and anchor graphs, incorporating view-specific spatial and spectral properties. Additionally, an adaptive view-weighting scheme is introduced to emphasize informative views and suppress less informative ones during the learning process. With the aid of index matrices $\{\mathbf{Q}^v\}_{v=1}^V$, anchor graph learning with multiple scales under incomplete settings is formulated as follows:

$$\min_{\mathbf{A}^v, \mathbf{Z}^v, \alpha} \sum_{v=1}^V \frac{1}{\alpha_v} \|\mathbf{X}^v \mathbf{Q}^v - \mathbf{A}^v \mathbf{Z}^v \mathbf{Q}^v\|_F^2, \quad (2)$$

$$\text{s.t. } \mathbf{Z}^v \geq 0, \mathbf{Z}^{v\top} \mathbf{1} = \mathbf{1}, \mathbf{A}^{v\top} \mathbf{A}^v = \mathbf{I}, \alpha^\top \mathbf{1} = 1, \alpha_v > 0,$$

where $\mathbf{Z}^v \mathbf{Q}^v$ represents the similarities between M_v anchors and N_v available pixels in the v -th view, and $\frac{1}{\alpha_v}$ measures the contribution of the v -th view. For effective learning, nonnegativity and normalization constraints are imposed on both $\mathbf{Z}^v$ and α , i.e., $\mathbf{Z}^v \geq 0$, $\mathbf{Z}^{v\top} \mathbf{1} = \mathbf{1}$, and $\alpha_v > 0$, $\alpha^\top \mathbf{1} = 1$, while an orthogonality constraint is applied to $\mathbf{A}^v$, enforcing $\mathbf{A}^{v\top} \mathbf{A}^v = \mathbf{I}$.

Structure-Aligned Consensus Feature Learning (SACFL). An intuitive method to incomplete multi-view clustering is to enforce consistency learning among anchor graphs across different views [13]. Due to inter-view heterogeneity, the learned view-specific anchor graphs exhibit varying scales to capture spatial and spectral characteristics, making it difficult to derive a unified graph representation. An alternative strategy is to learn a consensus feature representation [34]. However, significant structural inconsistencies across views hinder the extraction of consensus features in a shared latent space. To address this issue, alignment matrices $\{\mathbf{R}^v \in \mathbb{R}^{D \times M_v}\}_{v=1}^V$ are introduced to enable consensus feature learning by aligning the structures of multi-scale anchor graphs, with the orthogonality constraint $\mathbf{R}^{v\top} \mathbf{R}^v = \mathbf{I}$. Thus, consensus feature learning with structure alignment is mathematically formulated as follows:

$$\min_{\mathbf{R}^v, \mathbf{H}} \sum_{v=1}^V \|\mathbf{R}^v \mathbf{Z}^v - \mathbf{H}\|_F^2, \quad \text{s.t. } \mathbf{R}^{v\top} \mathbf{R}^v = \mathbf{I}, \quad (3)$$

where $\mathbf{H} \in \mathbb{R}^{D \times N}$ represents the consensus representations that aggregates discriminative information from multi-scale anchor graphs across all views.

Orthogonal Spatial-Aware Regularization (OSAR).

Neighboring pixels in remote sensing images often exhibit similar semantic information due to inherent spatial structures [42]. This property should be reflected in the learned consensus features by promoting continuity and smoothness in their spatial distribution. In the shared latent space, the continuous integral over the finite set of pixels can be discretized by computing finite differences between horizontally and vertically adjacent pixels, thereby enforcing local spatial consistency [9]. To this end, a total variation (TV)-based spatial-aware regularizer [4] is introduced to enhance spatial smoothness while preserving texture boundaries in both horizontal and vertical directions. Let $\mathcal{N}$ denote the set of neighboring pixels in the spatial domain. The regularizer is defined as follows:

$$\begin{aligned} \mathcal{R}(\mathbf{H}) &= \sum_{d=1}^D \sum_{\{i,j\} \in \mathcal{N}} \|\mathbf{h}_d^i - \mathbf{h}_d^j\|^2 \\ &= \sum_{d=1}^D \sum_{(i,j) \in \mathcal{N}_x} (\mathbf{h}_d^i - \mathbf{h}_d^j)^2 + \sum_{d=1}^D \sum_{(i,j) \in \mathcal{N}_y} (\mathbf{h}_d^i - \mathbf{h}_d^j)^2, \end{aligned} \quad (4)$$

where $\mathcal{N} = \mathcal{N}_x \cup \mathcal{N}_y$, with $\mathcal{N}_x$ and $\mathcal{N}_y$ being the horizontal and vertical neighbor sets. For simplicity, the orthogonal spatial-aware regularization in Eq. (4) can be rewritten as

$$\mathcal{R}(\mathbf{H}) = \|\mathbf{D}_x \mathbf{H}^\top\|_F^2 + \|\mathbf{D}_y \mathbf{H}^\top\|_F^2, \quad (5)$$

where $\|\mathbf{D}_x \mathbf{H}^\top\|_F^2$ and $\|\mathbf{D}_y \mathbf{H}^\top\|_F^2$ represent the horizontal and vertical differences, respectively, with $\mathbf{D}_x$ and $\mathbf{D}_y$ denoting the forward finite-difference operators in the horizontal and vertical directions [4].

Overall Objective Function. The separation between multi-scale anchor graph learning, structure-aligned consensus feature learning and orthogonal spatial-aware regularization may result in suboptimal performance. By combining Eqs. (2), (3) and (5), the objective function of OSMAGC is formulated as follows:

$$\begin{aligned} \min_{\substack{\mathbf{Z}^v, \mathbf{A}^v, \alpha \\ \mathbf{R}^v, \mathbf{H}}} \sum_{v=1}^V \frac{1}{\alpha_v} \|\mathbf{X}^v \mathbf{Q}^v - \mathbf{A}^v \mathbf{Z}^v \mathbf{Q}^v\|_F^2 &+ \lambda \sum_{v=1}^V \|\mathbf{R}^v \mathbf{Z}^v - \mathbf{H}\|_F^2 \\ &+ \gamma (\|\mathbf{D}_x \mathbf{H}^\top\|_F^2 + \|\mathbf{D}_y \mathbf{H}^\top\|_F^2), \\ \text{s.t. } \mathbf{Z}^v \geq 0, \mathbf{Z}^{v\top} \mathbf{1} = \mathbf{1}, \mathbf{A}^{v\top} \mathbf{A}^v = \mathbf{I}, \\ \mathbf{R}^{v\top} \mathbf{R}^v = \mathbf{I}, \alpha^\top \mathbf{1} = 1, \alpha_v > 0, \end{aligned} \quad (6)$$

where λ and γ represent trade-off parameters to balance each module. Through effective learning of each variable, the consensus representation $\mathbf{H}$ is refined and fed into k -means to yield the final clustering results.

Algorithm 1 The proposed OSMAGC

Require: Incomplete remote sensing data $\{\mathbf{X}^v\}_{v=1}^V$, super-pixel/anchor sizes $\{M^v\}_{v=1}^V$, index matrices $\{\mathbf{Q}^v\}_{v=1}^V$, parameters λ, γ , cluster number C .

- 1: Perform ERS segmentation on $\{\mathbf{X}^v\}_{v=1}^V$.
- 2: Generate $\{\mathbf{A}^v\}_{v=1}^V$, initialize $\{\mathbf{Z}^v\}_{v=1}^V$ using Eq. (1), initialize $\mathbf{H}$, and set $\{\alpha_v\}_{v=1}^V = 1/V$.
- 3: **while** not converged **do**
- 4: Update $\{\mathbf{A}^v\}_{v=1}^V$ by solving Eq. (8).
- 5: Update $\{\mathbf{R}^v\}_{v=1}^V$ by solving Eq.(9).
- 6: Update $\{\mathbf{Z}^v\}_{v=1}^V$ using Eq.(7).
- 7: Update $\mathbf{H}$ using Eq.(11).
- 8: Update $\{\alpha_v\}_{v=1}^V$ using Eq.(13).
- 9: **end while**

Ensure: Perform k -means on $\mathbf{H}$ to get the clustering results.

3.3. Optimization

To solve the problem in Eq. (6), an alternating optimization algorithm is designed to iteratively update each variable while keeping the others fixed.

Updating $\mathbf{Z}^v$: When $\mathbf{A}^v, \mathbf{R}^v, \mathbf{H}$ and α are fixed, the optimization for $\mathbf{Z}^v$ can be performed by independently updating each column vector $\mathbf{z}_j^v$ as

$$\min_{\mathbf{z}_j^v} \frac{1}{2} \|\mathbf{z}_j^v - \mathbf{f}_j^v\|_F^2, \quad \text{s.t.} \quad \mathbf{z}_j^v \geq 0, \mathbf{z}_j^{v\top} \mathbf{1} = 1, \quad (7)$$

where $f_{ij}^v = \frac{\frac{1}{\alpha_v} [\mathbf{A}^{v\top} \mathbf{X}^v \mathbf{S}^v]_{ij} + \lambda [\mathbf{R}^{v\top} \mathbf{H}]_{ij}}{\frac{1}{\alpha_v} s_{jj}^v + \lambda}$, with the diagonal matrix $\mathbf{S}^v = \mathbf{Q}^v \mathbf{Q}^{v\top}$, and $[\mathbf{R}^{v\top} \mathbf{H}]_{ij}$ denotes the element in the i -th row and j -th column of $\mathbf{R}^{v\top} \mathbf{H}$. Eq. (7) is a projection capped simplex problem. It can be efficiently solved as described in [29].

Updating $\mathbf{A}^v$: When $\mathbf{Z}^v, \mathbf{R}^v, \mathbf{H}$ and α are fixed, the optimization for $\mathbf{A}^v$ can be rewritten as

$$\max_{\mathbf{A}^v} \text{Tr}(\mathbf{A}^{v\top} \mathbf{M}^v), \quad \text{s.t.} \quad \mathbf{A}^{v\top} \mathbf{A}^v = \mathbf{I}, \quad (8)$$

where $\mathbf{M}^v = \mathbf{X}^v \mathbf{S}^v \mathbf{Z}^{v\top}$. According to [23], the optimal solution for Eq. (8) is given by $\mathbf{A}^v = \mathbf{U}^v \mathbf{V}^{v\top}$, where $\mathbf{U}^v$ and $\mathbf{V}^{v\top}$ are the matrices formed by the first M_v left singular vectors and right singular vectors of $\mathbf{M}^v$, respectively.

Updating $\mathbf{R}^v$: When $\mathbf{Z}^v, \mathbf{A}^v, \mathbf{H}$ and α are fixed, the optimization for $\mathbf{R}^v$ can be rewritten as

$$\max_{\mathbf{R}^v} \text{Tr}(\mathbf{R}^{v\top} \mathbf{B}^v), \quad \text{s.t.} \quad \mathbf{R}^{v\top} \mathbf{R}^v = \mathbf{I}, \quad (9)$$

where $\mathbf{B}^v = \mathbf{H} \mathbf{Z}^{v\top}$. The optimal $\mathbf{R}^v$ can be obtained by solving Eq. (9) with rank D truncated singular value decomposition (SVD) [23].

Updating $\mathbf{H}$: When $\mathbf{Z}^v, \mathbf{A}^v, \mathbf{R}^v$ and α are fixed, the

Table 1. Summary of the multi-view remote sensing datasets.

Datasets	Image Sizes	Clusters	View Details (Channels)
MUUF	325 × 220	11	HS(64), LiDAR(2)
Berlin	1723 × 476	8	HS(244), SAR(4)
Augsburg	332 × 485	7	HS(180), SAR(4), DSM(1)
MDAS	300 × 360	14	HS(242), MS(12), SAR(2), DSM(1)

optimization for $\mathbf{H}$ can be rewritten as

$$\min_{\mathbf{H}} -2\lambda \sum_{v=1}^V \text{Tr}(\mathbf{H}^\top \mathbf{R}^v \mathbf{Z}^v) + \lambda V \text{Tr}(\mathbf{H}^\top \mathbf{H}) + \gamma \text{Tr}(\mathbf{H} \mathbf{D} \mathbf{H}^\top), \quad (10)$$

where $\mathbf{D} = \mathbf{D}_x^\top \mathbf{D}_x + \mathbf{D}_y^\top \mathbf{D}_y$. Taking the derivative of $\mathbf{H}$ and setting it to zero, $\mathbf{H}$ is updated by

$$\mathbf{H} = \frac{\lambda \sum_{v=1}^V \mathbf{R}^v \mathbf{Z}^v}{\lambda V \mathbf{I} + \gamma \mathbf{D}}. \quad (11)$$

Updating α : When $\mathbf{Z}^v, \mathbf{A}^v, \mathbf{R}^v$ and $\mathbf{H}$ are fixed, the optimization for α can be rewritten as

$$\min_{\alpha} \sum_{v=1}^V \frac{(e_v)^2}{\alpha_v}, \quad \text{s.t.} \quad \alpha^\top \mathbf{1} = 1, \alpha_v > 0, \quad (12)$$

where $e_v = \|\mathbf{X}^v \mathbf{Q}^v - \mathbf{A}^v \mathbf{Z}^v \mathbf{Q}^v\|_F^2$. Based on the Cauchy-Schwarz inequality, each view weight in α is updated by

$$\alpha_v = \frac{e_v}{\sum_{v=1}^V e_v}. \quad (13)$$

Algorithm 1 summarizes the entire optimization process for solving Eq. (6).

3.4. Complexity Analysis

Space complexity: The memory consumption of OSMAGC primarily arises from storing the matrices $\mathbf{Z}^v \in \mathbb{R}^{M_v \times N}$, $\mathbf{A}^v \in \mathbb{R}^{B_v \times M_v}$, $\mathbf{R}^v \in \mathbb{R}^{D \times M_v}$ and $\mathbf{H} \in \mathbb{R}^{D \times N}$. Consequently, the overall space complexity is $\mathcal{O}(\sum_{v=1}^V (NM_v + M_v B_v + M_v D) + ND)$. Given that $M_v \ll N$, $B_v \ll N$ and $D \ll N$, the space complexity scales approximately linearly with the number of samples.

Time complexity: The computational complexity of OSMAGC mainly arises from the update of each variable. When updating $\mathbf{A}^v$, it takes $\mathcal{O}(NM_v B_v + M_v^2 B_v)$. For calculating $\mathbf{Z}^v$ and α_v , the complexity is $\mathcal{O}(NM_v B_v)$. The learning of $\mathbf{R}^v$ incurs a cost of $\mathcal{O}(NM_v D + M_v^2 D)$. The update of $\mathbf{H}$ requires $\mathcal{O}(DN_v + \sum_{v=1}^V NM_v D)$, where N_v denotes the number of nonzero elements in $\mathbf{D}$, and $N_v \ll N$. Therefore, the total time complexity per iteration is $\mathcal{O}(DN_v + \sum_{v=1}^V NM_v B_v + \sum_{v=1}^V NM_v D + \sum_{v=1}^V M_v^2 B_v + \sum_{v=1}^V M_v^2 D)$. This also indicates that the proposed optimization algorithm exhibits nearly linear time complexity with respect to the number of samples.

Table 2. Clustering performance of different methods on the four datasets with a 0.5 missing ratio. Best results are shown in **bold**, second-best are underlined, and “OOM” indicates out-of-memory.

Datasets	Metrics	FPFC TIP’24	AMKSC TNNLS’24	MSSAGF TGRS’25	SAMVGC AAAI’25	SIMVC-SA MM’23	DIVIDE AAAI’24	ASCR AAAI’25	PMIMC TIP’25	Ours
MUUFL	ACC	0.3410	0.3946	0.3913	0.3279	0.3999	0.3327	<u>0.4036</u>	0.3196	0.5002
	NMI	0.2705	0.3173	0.3101	0.2414	<u>0.3689</u>	0.2252	0.1689	0.2854	0.4024
	Purity	0.5698	0.6017	0.5738	0.5491	<u>0.6079</u>	0.5283	0.4573	0.5605	0.6639
	ARI	0.1067	<u>0.2019</u>	0.1191	0.1284	0.1818	0.1281	0.0799	0.1096	0.3351
Berlin	ACC	0.4010	0.3388	0.3501	0.3928	0.3649	0.3382	<u>0.4015</u>	OOM	0.4410
	NMI	0.1622	0.2140	0.2093	<u>0.2425</u>	0.1699	0.1701	0.0395	OOM	0.2534
	Purity	0.6401	0.6331	0.6474	<u>0.6850</u>	0.6424	0.6544	0.5781	OOM	0.7087
	ARI	0.0858	0.1385	0.1322	<u>0.1890</u>	0.0857	0.1147	0.0398	OOM	0.2098
Augsburg	ACC	<u>0.6009</u>	0.4102	0.4581	0.4550	0.4923	0.5083	0.4962	0.4104	0.6120
	NMI	<u>0.3815</u>	0.2616	0.3295	0.2811	0.3241	0.3302	0.1629	0.2660	0.4868
	Purity	0.7033	0.5482	0.6958	0.6679	<u>0.7240</u>	0.6845	0.5829	0.6191	0.8030
	ARI	<u>0.3418</u>	0.1705	0.2432	0.2285	0.2420	0.2498	0.1897	0.1709	0.3996
MDAS	ACC	<u>0.4171</u>	0.3282	0.3327	0.3804	0.3480	0.3064	0.3823	0.3292	0.4345
	NMI	0.2718	0.2841	0.2086	<u>0.3128</u>	0.2304	0.2653	0.1112	0.2508	0.3207
	Purity	0.5868	0.5990	0.5636	<u>0.6271</u>	0.5826	0.5789	0.5020	0.5450	0.6414
	ARI	<u>0.1850</u>	0.1680	0.0997	0.1552	0.1322	0.1032	0.1159	0.1052	0.1952

4. Experiments

4.1. Experimental Setup

Datasets. Four multi-view remote sensing image datasets are employed for the experiments, including MUUFL, Berlin, Augsburg and MDAS. Specifically, the MUUFL and Berlin datasets consist of two views, the Augsburg dataset includes three views, and the MDAS dataset comprises four views. These benchmarks exhibit progressively increasing complexity in terms of modality heterogeneity and spatial-spectral variation. Table 1 summarizes their key characteristics, including image size, view composition and the number of clusters. To construct incomplete datasets, a principled strategy is adopted to simulate real-world scenarios where different views provide complementary but partially missing information, with missing rates ranging from 0.1 to 0.9 in 0.1 intervals [29]. To ensure non-overlapping missing patterns across views, a sequential assignment scheme is applied. Each view is allocated a distinct set of missing pixels from the remaining pool. This ensures a globally accurate missing ratio and mutually exclusive patterns, providing a rigorous benchmark for evaluation.

Compared methods. Eight state-of-the-art methods are adopted for comprehensive comparison, including FPFC [40], AMKSC [2], MSSAGF [25], SAMVGC [8], SIMVC-SA [29], DIVIDE [17], ASCR [31] and PMIMC [35]. The first four baselines are multi-view clustering methods specifically designed for remote sensing data, whereas the latter four are incomplete multi-view clustering methods developed for general data.

Implementation details. For all compared methods, parameters are set according to their original papers to ensure the optimal performance. In our method, the number of superpixels is tuned from 100 to 300 in 50 intervals, and the regularization parameters λ and γ are selected from [0.001, 0.01, 0.1, 1, 10, 100, 1000]. Clustering performance is evaluated using accuracy (ACC), normalized mutual information (NMI), purity, and adjusted rand index (ARI). All experiments are conducted on a server with an Intel Xeon Gold 6330 2.00GHz CPU, 256GB RAM, and two NVIDIA GeForce RTX 4090 24GB GPUs. To reduce randomness, each method is executed ten times using the optimal parameter configurations for each missing rate, and the average results are reported.

4.2. Comparison Results

Clustering Performance. Table 2 shows the clustering results of different methods under a 0.5 missing ratio. Our method consistently outperforms all competitors across all datasets, achieving superior results in terms of the four metrics. In particular, our method yields ACC improvements of 9.66%, 3.95%, 1.11%, and 1.74% on the four datasets, respectively. Furthermore, Figure 3 illustrates the clustering results across a range of missing ratios from 0.1 to 0.9 in 0.1 increments. At a missing ratio of 0.3, our method outperforms the second-best method by 5.00%, 3.61%, 4.63% and 11.64% in the four metrics on the MUUFL dataset, respectively. When the missing ratio increases to 0.7, OSMAGC still achieves ACC improvements of 6.78%, 2.82%, 3.28% and 2.86% across the four datasets. Similar performance advantages can be consistently observed at other missing ratios,

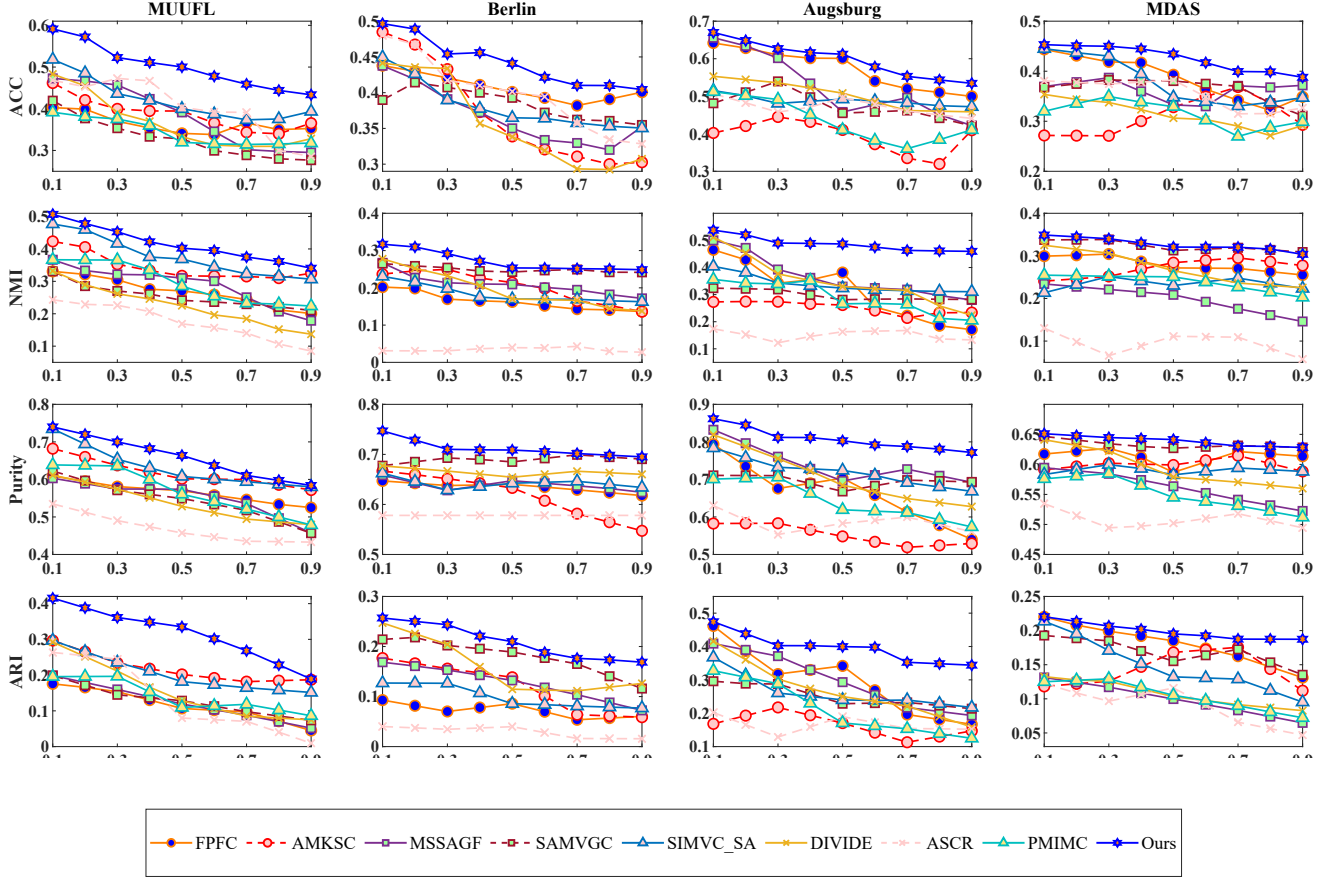


Figure 3. Clustering performance of different methods on the four datasets across different missing ratios.

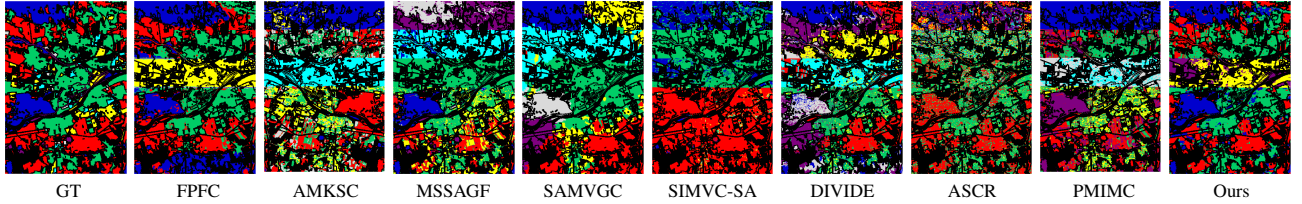


Figure 4. Clustering maps of the Augsburg dataset obtained by different methods under a 0.5 missing ratio. GT: Ground truth.

demonstrating the robustness and generalization capability of our method. The superior performance of OSMAGC can be primarily attributed to its ability to effectively recover missing information in a latent space by exploiting cross-view complementarity and consistency, while simultaneously preserving spatial and spectral continuity.

To visualize the clustering effect, Figure 4 presents the clustering maps of the Augsburg dataset produced by different methods under a 0.5 missing ratio. Our method achieves the closest alignment with the ground truth by effectively recovering missing information while preserving spatial structure, boundary clarity and local details. In contrast, other competitors struggle to handle remote sensing data with

missing information, resulting in pronounced fragmented regions, category confusion and blurred boundaries. This is because the first four methods have limited capability in handling incomplete data with missing information, whereas the latter four fail to capture the inherent spatial-spectral structures of remote sensing data. These observations further highlight the robustness and superiority of our method in clustering incomplete multi-view remote sensing data.

Running Time. Table 3 reports the running time of different methods under a 0.5 missing ratio. As shown, DIVIDE and PMIMC are highly time-consuming when partitioning the four datasets, and PMIMC even runs out of memory on the Berlin dataset. In contrast, the proposed OSMAGC

Table 3. Running time of different methods on the four datasets under a 0.5 missing ratio (in seconds).

Methods	MUUFL	Berlin	Augsburg	MDAS
FPFC	<u>16.19</u>	127.48	39.1	<u>20.68</u>
AMKSC	106.81	742.75	177.86	213.58
MSSAGF	48.7	599.2	133.03	140.02
SAMVGC	145.07	145.92	52.56	169.79
SIMVC-SA	17.09	200.02	<u>37.26</u>	24.57
DIVIDE	1735.34	1809.31	2548.46	1277.68
ASCR	16.92	<u>71.46</u>	45.46	75.86
PMIMC	1890.5	OOM	19744.62	20084.99
Ours	14.01	68.38	31.6	14.78

Table 4. Ablation study of our OSMAGC on the four datasets under a 0.5 missing ratio (in ACC).

Variants	MUUFL	Berlin	Augsburg	MDAS
V1 (w/o SAGI)	0.4591	0.4129	0.5325	0.3220
V2 (w/o MSAGL)	0.4713	0.4281	0.5530	0.4240
V3 (w/o SACFL)	0.4463	0.4355	0.5249	0.3360
V4 (w/o OSAR)	0.4661	0.4375	0.5689	0.4211
Ours	0.5002	0.4410	0.6120	0.4345

achieves the highest efficiency, requiring only 14.01, 68.38, 31.6 and 14.78 seconds to cluster the four datasets. Compared with the second fastest methods, it reduces execution time by 2.18, 3.08, 5.66 and 5.9 seconds on the four datasets, respectively. These observations highlight the practical potential of our method for real-world applications.

4.3. Model Analysis

Ablation Study. To validate the effectiveness of each component, four degraded variants are constructed. Specifically, “V1” removes SAGI and uses k-means to generate anchors, “V2” removes MSAGL and assigns a same superpixel size to each view, “V3” removes SACFL by keeping the alignment matrix fixed, and “V4” removes OSAR. As shown in Table 4, our method consistently outperforms all degraded variants across datasets under a 0.5 missing ratio, confirming the effectiveness of each component. For instance, removing SAGI in “V1” leads to the largest performance drop on the MDAS dataset, while eliminating MSAGL in “V2” most markedly affects the Augsburg dataset. The MUUFL dataset is most sensitive to the removal of SACFL in “V3”, whereas the Berlin dataset exhibits only a minor decline when OSAR is omitted in “V4”. These dataset-specific differences highlight the adaptive importance of each component. Overall, the consistent superiority of our full model demonstrates the synergistic effect and necessity of all components.

Parameter Analysis. Our method involves three key parameters: superpixel size and two regularization parameters, λ and γ . Due to the page limit, only the results on

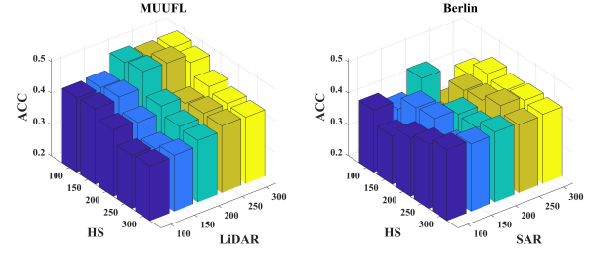


Figure 5. Performance variation of our OSMAGC under different superpixel settings with a 0.5 missing ratio.

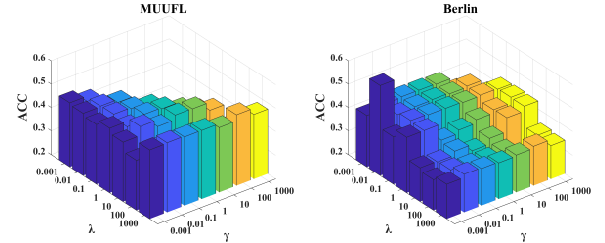


Figure 6. Performance variation of our OSMAGC under different settings of λ and γ with a 0.5 missing ratio.

the MUUFL and Berlin datasets are presented. As shown in Figure 5, the number of superpixels affects clustering performance differently across datasets. The MUUFL dataset achieves stable and slightly improved results with more superpixels, while the Berlin dataset performs better with fewer superpixels and declines when the number is too high. In addition, Figure 6 shows the sensitivity of λ and γ . Both parameters are stable on the MUUFL dataset, whereas the Berlin dataset is more sensitive to λ , with excessive values causing performance drops. Optimal results for both datasets are obtained when γ is within a moderate range, highlighting the importance of appropriate parameter settings.

5. Conclusion

This paper proposes an orthogonal spatial-aware multi-view anchor graph clustering (OSMAGC) method that first investigates incomplete remote sensing data. OSMAGC captures diverse spatial characteristics across views, enabling the initialization of multi-scale anchor graphs for more effective subsequent learning. A unified objective function is formulated by integrating multi-scale anchor graph learning, structure-aligned consensus feature learning and an orthogonal spatial-aware regularizer. An alternating optimization algorithm is devised for the proposed formulation. OSMAGC effectively exploits complementary and consistent information across views while preserving spatial and spectral continuity. Extensive experiments demonstrate the superiority of our proposed method.

6. Acknowledgments

This work was funded by the China National Key R&D Program (Grant 2023YFF0807000).

References

- [1] Yaoming Cai, Zijia Zhang, Pedram Ghamisi, Behnood Rasti, Xiaobo Liu, and Zhihua Cai. Transformer-based contrastive prototypical clustering for multimodal remote sensing data. *Information Sciences*, 649:119655, 2023. 1
- [2] Yaoming Cai, Zijia Zhang, Xiaobo Liu, Yao Ding, Fei Li, and Jinhua Tan. Learning unified anchor graph for joint clustering of hyperspectral and lidar data. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4):6341–6354, 2025. 1, 2, 6
- [3] Zhe Cao, Haonan Xin, Rong Wang, and Feiping Nie. Superpixel-based bipartite graph clustering enriched with spatial information for hyperspectral and lidar data. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025. 2
- [4] Yi Chang, Luxin Yan, Houzhang Fang, and Chunan Luo. Anisotropic spectral-spatial total variation model for multi-spectral remote sensing image destriping. *IEEE Transactions on Image Processing*, 24(6):1852–1866, 2015. 4
- [5] Zhengyi Chen, Chunmin Zhang, Tingkui Mu, and Yifan He. Tensorial multiview subspace clustering for polarimetric hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. 2
- [6] Zhibin Dong, Jiaqi Jin, Yuyang Xiao, Bin Xiao, Siwei Wang, Xinwang Liu, and En Zhu. Subgraph propagation and contrastive calibration for incomplete multiview data clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. 2
- [7] Renxiang Guan, Zihao Li, Wenxuan Tu, Jun Wang, Yue Liu, Xianju Li, Chang Tang, and Ruyi Feng. Contrastive multi-view subspace clustering of hyperspectral images based on graph convolutional networks. *IEEE Transactions on Geoscience and Remote Sensing*, 62:5510514, 2024. 1, 2
- [8] Renxiang Guan, Wenxuan Tu, Siwei Wang, Jiyuan Liu, Dayu Hu, Chang Tang, Yu Feng, Junhong Li, Baili Xiao, and Xinwang Liu. Structure-adaptive multi-view graph clustering for remote sensing data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 16933–16941, 2025. 1, 2, 6
- [9] Chengcai Leng, Hai Zhang, Guorong Cai, Zhen Chen, and Anup Basu. Total variation constrained non-negative matrix factorization for medical image registration. *IEEE/CAA Journal of Automatica Sinica*, 8(5):1025–1037, 2021. 4
- [10] Chengliang Liu, Zhihao Wu, Jie Wen, Yong Xu, and Chao Huang. Localized sparse incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 25:5539–5551, 2022. 2
- [11] Jianlun Liu, Shaohua Teng, Wei Zhang, Xiaozhao Fang, Lunke Fei, and Zhuxiu Zhang. Incomplete multi-view subspace clustering with low-rank tensor. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3180–3184. IEEE, 2021. 2
- [12] Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam, and Rama Chellappa. Entropy rate superpixel segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2097–2104. IEEE, 2011. 3
- [13] Suyuan Liu, Qing Liao, Siwei Wang, Xinwang Liu, and En Zhu. Robust and consistent anchor graph learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36(8):4207–4219, 2024. 4
- [14] Suyuan Liu, Siwei Wang, Ke Liang, Junpu Zhang, Zhibin Dong, Tianrui Liu, En Zhu, Xinwang Liu, and Kunlun He. Alleviate anchor-shift: Explore blind spots with cross-view reconstruction for incomplete multi-view clustering. *Advances in Neural Information Processing Systems*, 37:87509–87531, 2024. 2
- [15] Xinwang Liu, Xinzong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang, and Wen Gao. Late fusion incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2410–2423, 2018. 2
- [16] Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft, and En Zhu. Efficient and effective regularized incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2634–2646, 2020. 2
- [17] Yiding Lu, Yijie Lin, Mouxiang Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 14193–14201, 2024. 6
- [18] Fulin Luo, Yi Liu, Yule Duan, Tan Guo, Lefei Zhang, and Bo Du. Sdst: Self-supervised double-structure transformer for hyperspectral images clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 62:5511614, 2024. 2
- [19] Yalan Qin, Guorui Feng, and Xinpeng Zhang. Scalable one-pass incomplete multi-view clustering by aligning anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 20042–20050, 2025. 4
- [20] Huanfeng Shen, Xinghua Li, Qing Cheng, Chao Zeng, Gang Yang, Huifang Li, and Liangpei Zhang. Missing information reconstruction of remote sensing data: A technical review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):61–85, 2015. 2
- [21] Yuting Wan, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. Multiobjective sine cosine algorithm for remote sensing image spatial-spectral clustering. *IEEE Transactions on Cybernetics*, 52(10):11172–11186, 2021. 2
- [22] Huibing Wang, Mingze Yao, Yawei Chen, Yunqiu Xu, Haipeng Liu, Wei Jia, Xianping Fu, and Yang Wang. Manifold-based incomplete multi-view clustering via bi-consistency guidance. *IEEE Transactions on Multimedia*, 26:10001–10014, 2024. 2
- [23] Siwei Wang, Xinwang Liu, En Zhu, Chang Tang, Jiyuan Liu, Jingtao Hu, Jingyuan Xia, and Jianping Yin. Multi-view clustering via late fusion alignment maximization. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3778–3784, 2019. 5
- [24] Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering

- with adaptive anchor fusion. *IEEE Transactions on Image Processing*, 33:4627–4639, 2024. 4
- [25] Xinxin Wang, Yongshan Zhang, and Yicong Zhou. Multimodal remote sensing image clustering with multi-scale spectral-spatial anchor graphs. *IEEE Transactions on Geoscience and Remote Sensing*, 63:4405612, 2025. 1, 2, 6
- [26] Jie Wen, Zheng Zhang, Zhao Zhang, Zhihao Wu, Lunke Fei, Yong Xu, and Bob Zhang. Dimc-net: Deep incomplete multi-view clustering network. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3753–3761, 2020. 2
- [27] Jie Wen, Chengliang Liu, Gehui Xu, Zhihao Wu, Chao Huang, Lunke Fei, and Yong Xu. Highly confident local structure based consensus graph learning for incomplete multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15712–15721, 2023. 2
- [28] Jie Wen, Gehui Xu, Zhanyan Tang, Wei Wang, Lunke Fei, and Yong Xu. Graph regularized and feature aware matrix factorization for robust incomplete multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):3728–3741, 2023. 2
- [29] Yi Wen, Siwei Wang, Ke Liang, Weixuan Liang, Xinhang Wan, Xinwang Liu, Suyuan Liu, Jiyan Liu, and En Zhu. Scalable incomplete multi-view clustering with structure alignment. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3031–3040, 2023. 2, 5, 6
- [30] Danyang Wu, Zhenkun Yang, Jitao Lu, Jin Xu, Xiangmin Xu, and Feiping Nie. Ebmgc-gnf: Efficient balanced multi-view graph clustering via good neighbor fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):7878–7892, 2024. 1
- [31] Deng Xu, Chao Zhang, Cong Guo, Chunlin Chen, and Huaxiong Li. Fast incomplete multi-view clustering with adaptive similarity completion and reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 21734–21742, 2025. 1, 6
- [32] Jun Yin and Shiliang Sun. Incomplete multi-view clustering with reconstructed views. *IEEE Transactions on Knowledge and Data Engineering*, 35(3):2671–2682, 2021. 2
- [33] Jun Yin and Shiliang Sun. Incomplete multi-view clustering with cosine similarity. *Pattern Recognition*, 123:108371, 2022. 2
- [34] Xiao Yu, Hui Liu, Yan Zhang, Shanbao Sun, and Caiming Zhang. Multi-view clustering via efficient representation learning with anchors. *Pattern Recognition*, 144:109860, 2023. 4
- [35] Honglin Yuan, Yuan Sun, Fei Zhou, Jing Wen, Shihua Yuan, Xiaojian You, and Zhenwen Ren. Prototype matching learning for incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 2025. 1, 6
- [36] Lingbin Zeng, Shixin Yao, You Huang, Yong Cheng, and Yue Qian. Improved multi-view graph clustering with global graph refinement. *Remote Sensing*, 17(18):3217, 2025. 2
- [37] Biao Zhang, Mengmeng Zhang, Yuxiang Zhang, Huan Liu, Zhengqi Guo, Xiaoming Xie, and Wei Li. Hyperspectral and multi-source heterogeneous data fusion classification based on multiscale multi-source interaction attention network. In *Proceedings of the 2024 7th International Conference on Image and Graphics Processing*, pages 217–223, 2024. 2
- [38] Lefei Zhang, Qian Zhang, Bo Du, Xin Huang, Yuan Yan Tang, and Dacheng Tao. Simultaneous spectral-spatial feature selection and extraction for hyperspectral images. *IEEE Transactions on Cybernetics*, 48(1):16–28, 2016. 1
- [39] Yongshan Zhang, Shuaikang Yan, Xinwei Jiang, Lefei Zhang, Zhihua Cai, and Jun Li. Dual graph learning affinity propagation for multimodal remote sensing image clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 1, 2
- [40] Yongshan Zhang, Shuaikang Yan, Lefei Zhang, and Bo Du. Fast projected fuzzy clustering with anchor guidance for multimodal remote sensing imagery. *IEEE Transactions on Image Processing*, 33:4640–4653, 2024. 1, 2, 3, 6
- [41] Yingying Zhang, Lixiang Ru, Kang Wu, Lei Yu, Lei Liang, Yansheng Li, and Jingdong Chen. Skysense v2: A unified foundation model for multi-modal remote sensing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9136–9146, 2025. 1
- [42] Hao Zhu, Yan Zhu, Jiayu Xiao, Tianxiang Xiao, Yike Ma, Yucheng Zhang, and Feng Dai. Exact: Exploring space-time perceptive clues for weakly supervised satellite image time series semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14036–14045, 2025. 4