

# Efficient Tensorized Multi-View Anchor Graph Clustering with Affinity Propagation for Remote Sensing Data

Yongshan Zhang<sup>1</sup>, Kangyue Zheng<sup>1</sup>, Shuaikang Yan<sup>1</sup>, Xinxin Wang<sup>2\*</sup>, Zhihua Cai<sup>1</sup>

<sup>1</sup>School of Computer Science, China University of Geosciences, Wuhan, China

<sup>2</sup>School of Artificial Intelligence, Shenzhen University, Shenzhen, China  
 yzhang.cug@gmail.com, zhengkangyue@cug.edu.cn, 1202221621ysk@cug.edu.cn,  
 xinxinwang1024@gmail.com, zhcai@cug.edu.cn

## Abstract

Multi-view clustering of remote sensing data presents significant challenges, as it integrates diverse data representations to improve Earth observation. Although existing anchor graph-based methods have yielded promising results, they generally exhibit two key limitations: (1) the time-consuming process of directly exploring pixel clustering structures, and (2) insufficient modeling of high-order correlations among different views. To address these issues, we propose an **Efficient Tensorized multi-view anchor graph clustering method with Affinity Propagation (ETAP)** for remote sensing data. Based on superpixel preprocessing, anchor graphs are learned from view-specific pixels and anchors, while compressed anchor graphs are simultaneously learned from the view-specific anchors. An adaptive weighting scheme is introduced to facilitate the learning of these anchor graphs. To capture high-order correlations, tensor Schatten  $p$ -norm regularization is applied to the compressed anchor graphs. A connectivity constraint is introduced to uncover the clustering structures of anchors. Finally, pixel clustering structures are then efficiently revealed from the pseudo-labeled anchors through affinity propagation without requiring additional clustering steps. To solve the proposed formulation, we develop an alternating optimization algorithm. Extensive experiments on three public datasets demonstrate the efficacy and efficiency of the proposed method over state-of-the-art methods.

**Code** — <https://github.com/ZhangYongshan/ETAP>

## Introduction

Remote sensing data can be collected from various perspectives, providing multiple representations to enhance Earth observation (Yang, Li, and Zhang 2023; Zhang and Zhang 2022; Yang et al. 2023). The emergence of multi-view remote sensing data requires multi-view learning techniques to effectively integrate complementary information for better scene understanding (Xue et al. 2022; Zhang et al. 2020, 2023). Due to the time-consuming and labor-intensive nature of large-scale pixel labeling, multi-view clustering for remote sensing data has gained significant attention by leveraging cross-view consistency and complementarity to partition pixels into distinct clusters without the need for ground-

truth labels (Wang et al. 2024). It has been successfully applied in various fields, including precision agriculture, environmental protection and military defense.

Recently, a large number of multi-view clustering methods have been specifically proposed for remote sensing data. Based on the features utilized in different views, these methods can be divided into multi-feature-based and multi-source-based methods (Guan et al. 2025). Multi-feature-based methods address the clustering problem of single-sensor remote sensing data represented by multi-view features, such as texture and contour (Luo et al. 2024). However, they are inadequate for capturing the complementary information in multi-view data derived from multiple sensors (Cai et al. 2025; Zhang et al. 2024a). To solve this issue, multi-source-based methods tackle the clustering problem of multi-view remote sensing data collected by different sensors, such as hyperspectral (HS), multispectral (MS) and synthetic aperture radar (SAR) images (Zhang et al. 2024b; Guan et al. 2025). Compared to single-sensor data, the exploration of multi-sensor data is more challenging due to the complex spatial distribution and significant spectral variations across different views (Li et al. 2024b).

Anchor graph-based clustering has emerged as an effective and efficient technique that explores the relationships between pixels and representative anchors to analyze multi-view remote sensing data captured by different sensors (Cai et al. 2025; Zhang et al. 2024a,b; Li et al. 2025). Based on the clustering process, anchor graph-based methods are primarily categorized into post-processing and one-pass methods. Post-processing methods require an additional  $k$ -means clustering step applied to the fused anchor graph for pixel-wise partitioning (Sun et al. 2022), whereas one-pass methods directly uncover pixel clustering structures by imposing connectivity constraints or learning clustering indicators (Liu et al. 2024). Although existing methods demonstrate satisfactory results, they primarily suffer from two limitations: (1) most of them are limited by the time-consuming process of directly exploring the clustering structures of pixels; and (2) many of them exhibit insufficient modeling of high-order correlations across different views.

To overcome the above limitations, in this paper, we propose an **Efficient Tensorized multi-view anchor graph clustering method with Affinity Propagation (ETAP)** for remote sensing data, as shown in Figure 1. Specifically, nonlin-

\*Corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

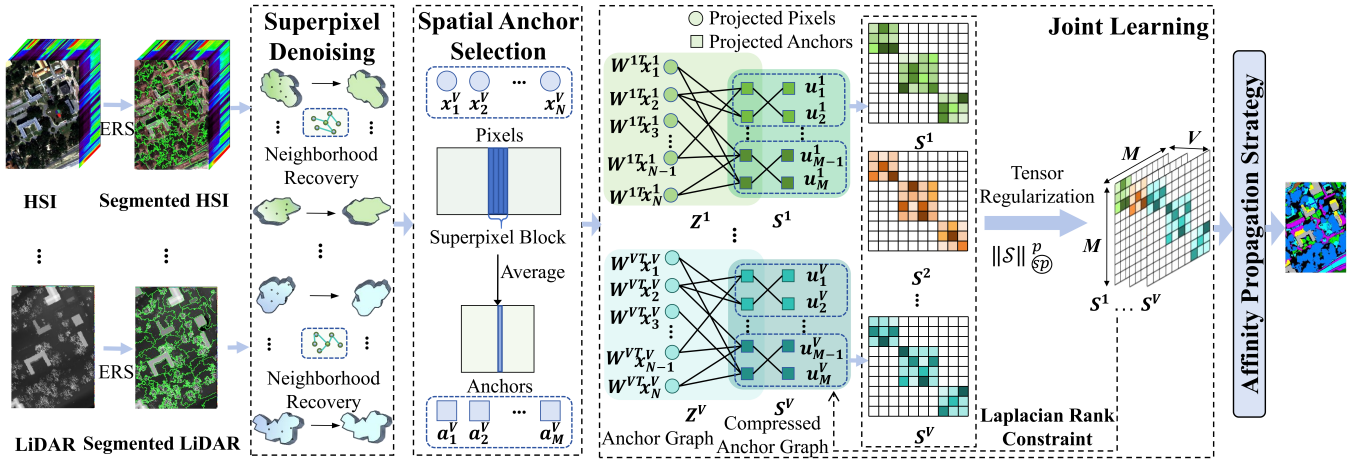


Figure 1: Illustration of the proposed ETAP.

ear denoising and anchor initialization are first performed based on the spatial texture characteristics of each view. After spatial preprocessing, anchor graphs and compressed anchor graphs are simultaneously learned using an adaptive weighting scheme to capture the relationships between view-specific pixels and anchors, as well as among view-specific anchors. To accelerate the learning process, tensor Schatten  $p$ -norm regularization and a connectivity constraint are applied to the compressed anchor graphs rather than the original anchor graphs, enabling the modeling of high-order correlations and the discovery of anchor clustering structures. Pixel clustering structures can be efficiently revealed from the pseudo-labeled anchors based on affinity propagation. An alternating optimization algorithm is developed to solve the proposed formulation. Extensive experiments demonstrate the efficacy and efficiency of the proposed method over state-of-the-art competitors.

In summary, the main contributions of this paper are highlighted as follows.

- We propose a novel multi-view anchor graph clustering framework that simultaneously learns view-specific anchor graphs and compressed anchor graphs to efficiently explore the clustering structures of multi-sensor remote sensing data.
- We explore high-order correlations among views using tensor Schatten  $p$ -norm regularization and uncover the clustering structures of anchors with a connectivity constraint, both applied to the compressed anchor graphs to accelerate the learning process.
- We design an efficient affinity propagation strategy to assign clustering labels to multi-view pixels from the pseudo-labeled anchors in the structured compressed anchor graphs.
- An alternating optimization algorithm is developed to solve the proposed formulation. Extensive experiments demonstrate the superior efficacy and efficiency of our proposed method.

## Related Work

Due to the diversity of data representations and the absence of ground-truth labels, various multi-view clustering methods have been proposed to achieve pixel partitioning for remote sensing data (Cai et al. 2022; Wan et al. 2022; Wang and Zhu 2020). They can group pixels into distinct clusters by leveraging multi-view information without label guidance. Based on the nature of multi-view information, these methods can be broadly separated into two categories: multi-feature-based and multi-sensor-based methods. For multi-feature-based methods, they perform multi-view clustering on multiple features derived from single-sensor remote sensing data. For instance, Luo et al. (Luo et al. 2024) utilize a Transformer with self-supervised learning to fuse spatial and spectral views for HS image clustering. Similarly, Chen et al. (Chen et al. 2022) perform polarimetric HS image clustering by organizing views into a tensor and imposing consistency constraints to strengthen inter-view correlations. For multi-sensor-based methods, they integrate multi-view remote sensing data from different sensors for clustering. Specifically, Zhang et al. (Zhang et al. 2024b) learn a shared anchor graph and fuzzy membership matrix from multi-sensor data to facilitate clustering. Further, Guan et al. (Guan et al. 2025) adaptively refine graph and cluster structures through contrastive and consistent learning for multi-sensor data.

The aforementioned methods demonstrate promising performance in clustering multi-sensor remote sensing data. However, they often involve time-consuming processes to directly explore pixel clustering structures and fail to effectively model high-order correlations across different views.

## Preliminary

For clarity, notations and definitions are introduced here. Scalars, vectors, matrices and tensors are denoted by lowercase letters, boldface lowercase letters, boldface uppercase letters and boldface calligraphic letters, respectively. For a matrix  $\mathbf{S}$ , we denote its trace, transpose and  $F$ -norm as  $\text{tr}(\mathbf{S})$ ,  $\mathbf{S}^T$  and  $\|\mathbf{S}\|_F$ , respectively. For a third-order ten-

tensor  $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , the  $i$ -th frontal slice is denoted as  $\mathcal{S}^{(i)}$ . The discrete Fourier transform (DFT) of  $\mathcal{S}$  along the third dimension is given by  $\bar{\mathcal{S}} = \text{fft}(\mathcal{S}, [], 3)$ , and its reverse operation is  $\mathcal{S} = \text{ifft}(\bar{\mathcal{S}}, [], 3)$ . Due to space limit, we provide only the definition of the tensor Schatten  $p$ -norm. More related definitions can be found in (Kilmer and Martin 2011).

**Definition 1.** (Gao et al. 2021) Given  $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ ,  $h = \min(n_1, n_2)$ , the tensor Schatten  $p$ -norm of  $\mathcal{S}$  is defined as

$$\|\mathcal{S}\|_{\mathbb{S}}^p = \left( \sum_{i=1}^{n_3} \|\bar{\mathcal{S}}^{(i)}\|_{\mathbb{S}}^p \right)^{\frac{1}{p}} = \left( \sum_{i=1}^{n_3} \sum_{j=1}^h \sigma_j(\bar{\mathcal{S}}^{(i)})^p \right)^{\frac{1}{p}}, \quad (1)$$

where  $\sigma_j(\bar{\mathcal{S}}^{(i)})$  is the  $j$ -th singular value of matrix  $\bar{\mathcal{S}}^{(i)}$ . It should be noted that when  $p$  is appropriately chosen, the Schatten  $p$ -norm effectively improves the approximation of the rank function than other norms (Li et al. 2024a).

## Method

### Formulation

**Superpixel Denoising and Anchor Initialization** To mitigate the noise incurred by challenging observation conditions, denoising can be guided by spatial information of remote sensing data (Lu et al. 2023; Shi, Huang, and Wang 2020). Given a set of remote sensing images  $\{\mathcal{X}^v\}_{v=1}^V$  with  $V$  views,  $\mathcal{X}^v \in \mathbb{R}^{W \times H \times B_v}$  represents the  $v$ -th view image, containing  $W \times H$  pixels and  $B_v$  spectral channels. To generate  $M$  view-specific superpixels that reflect the spatial distribution and texture complexity, we apply the entropy rate superpixel (ERS) segmentation method (Liu et al. 2011) to the first principal component of each image, owing to its promising efficacy and efficiency. Based on this, a superpixel-based reconstruction method is introduced for denoising. It follows that all original pixels within the same superpixel are close to the recovered ones and are more likely to be distributed on nonlinear geometric manifolds (Wang, Zhang, and Zhou 2025). Assume that the manifold structure of all pixels within a superpixel is represented by  $\mathbf{M} \in \mathbb{R}^{L \times L}$ , where  $L$  denotes the number of pixels in the superpixel. The original pixels are then reconstructed as follows:

$$\min_{\hat{\mathbf{Y}}} \|\hat{\mathbf{Y}} - \mathbf{Y}\|_F^2 + \xi \text{Tr}(\mathbf{Y} \mathbf{L}_M \mathbf{Y}^T), \quad (2)$$

where  $\hat{\mathbf{Y}} \in \mathbb{R}^{B_v \times L}$  indicates the  $L$  observed pixels within a superpixel, and  $\mathbf{Y} \in \mathbb{R}^{B_v \times L}$  denotes the corresponding recovered pixels. The parameter  $\xi$  is used to balance the distance deviation and manifold constraint, and  $\mathbf{L}_M = \mathbf{D}_M - \mathbf{M}$  is a Laplacian matrix, where  $\mathbf{D}_M$  is a degree matrix with diagonal elements  $(d_M)_{ii} = \sum_{j=1}^L m_{ij}$ . After denoising, the reconstructed remote sensing images are organized in matrix form as  $\{\mathbf{X}^v\}_{v=1}^V$ , with each image in the  $v$ -th view represented as  $\mathbf{X}^v = \{\mathbf{x}_1^v, \dots, \mathbf{x}_N^v\} \in \mathbb{R}^{B_v \times N}$ , where  $N = H \times W$  denotes the pixel size.

High-quality anchors are crucial for effective anchor graph-based clustering. To this end, we also generate high-quality anchors based on the generated superpixels, specifically tailored for remote sensing data. For instance, consider a superpixel with  $L$  pixels, the corresponding anchor

$\mathbf{a} \in \mathbb{R}^{B_v \times 1}$  is generated by averaging all reconstructed pixels within the superpixel. In this way, we can obtain multi-view spatial-aware anchors  $\{\mathbf{A}^v\}_{v=1}^V$ , where  $\mathbf{A}^v = \{\mathbf{a}_1^v, \dots, \mathbf{a}_M^v\} \in \mathbb{R}^{B_v \times M}$  corresponds to  $M$  anchors in the  $v$ -th view. This anchor graph initialization method is able to uncover pixel distributions and preserve spatial-spectral discriminative information, enabling the generation of high-quality anchors for effective subsequent learning.

**Anchor Graph Learning** Compared to traditional graphs that describe the relationships between pairwise samples, anchor graphs reduce their size by representing the relationships between samples and representative anchors (Wang et al. 2022). This results in higher efficiency, significantly reducing both computational complexity and memory consumption for large-scale problems. Given  $\{\mathbf{X}^v\}_{v=1}^V$  and  $\{\mathbf{A}^v\}_{v=1}^V$ , we adaptively learn the anchor graphs  $\{\mathbf{Z}^v\}_{v=1}^V$  through projection learning for different views. Here,  $\mathbf{Z}^v \in \mathbb{R}^{N \times M}$  represents the anchor graph for the  $v$ -th view, where  $z_{ij}^v \in \mathbf{Z}^v$  encodes the relationships between the  $i$ -th pixel and  $j$ -th anchor. Since the information varies across different views, an auto-weighted scheme is introduced to adaptively quantify the contributions of each view. To learn  $\mathbf{Z}^v$ , pixel  $\mathbf{x}_i^v$  should be connected to anchor  $\mathbf{a}_j^v$  with a probability  $z_{ij}^v$  to describe their neighboring relationship. The larger value of  $z_{ij}^v$  corresponds to a shorter distance between  $\mathbf{x}_i^v$  and  $\mathbf{a}_j^v$ , and vice versa. Thus, anchor graph learning for multi-view remote sensing data can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{W}^v, \mathbf{Z}^v, \alpha^v} & \sum_{v=1}^V \frac{1}{\alpha^v} \sum_{i=1}^N \sum_{j=1}^M \left\| \mathbf{W}^{vT} \mathbf{x}_i^v - \mathbf{a}_j^v \right\|_2^2 z_{ij}^v + \delta \sum_{v=1}^V \|\mathbf{Z}^v\|_F^2 \\ \text{s.t.} & \quad \mathbf{z}_i^{vT} \mathbf{1} = 1, z_{ij}^v \geq 0, \boldsymbol{\alpha}^T \mathbf{1} = 1, \alpha^v > 0, \\ & \quad \mathbf{W}^{vT} \mathbf{S}_t^v \mathbf{W}^v = \mathbf{I}. \end{aligned} \quad (3)$$

where  $\mathbf{W}^v \in \mathbb{R}^{B_v \times R_v}$  is the projection matrix used to perform learning in a  $R_v$ -dimensional reduced subspace,  $\mathbf{U}^v = \mathbf{W}^{vT} \mathbf{A}^v \in \mathbb{R}^{R_v \times M}$  is the projected anchor matrix for easy representation,  $\delta$  represents a regularization parameter, and  $\frac{1}{\alpha^v}$  denotes the view weight. Note that superscript or subscript  $v$  refers to the  $v$ -th view. For effective learning, nonnegativity and normalization constraints are imposed on both  $\mathbf{Z}^v$  and  $\alpha^v$  (Wang et al. 2022, 2019), while the constraint  $\mathbf{W}^{vT} \mathbf{S}_t^v \mathbf{W}^v = \mathbf{I}$  ensures that the projected data are statistically uncorrelated, where  $\mathbf{S}_t^v = \mathbf{X}^v \mathbf{X}^{vT}$  represents the total scatter matrix.

**Optimal Compressed Anchor Graph Learning** When the sample size is large, anchor graph-based methods still require a time-consuming process to learn the exact connected components or necessitate post-processing clustering steps (Cai et al. 2025). To solve this problem, we learn the compressed anchor graphs  $\{\mathbf{S}^v\}_{v=1}^V$  with connectivity constraints from  $\{\mathbf{U}^v\}_{v=1}^V$  in the reduced subspace. Here,  $\mathbf{S}^v \in \mathbb{R}^{M \times M}$  represents the compressed anchor graph for the  $v$ -th view, where  $s_{ij}^v \in \mathbf{S}^v$  encodes the relationships between pair-wise anchors. Besides, high-order correlations among multi-view data are often overlooked in remote sensing image analysis. To capture these multi-way dependencies, we construct a tensor by stacking the multi-view compressed anchor graphs and impose a tensor Schatten- $p$  norm regularization. This regularizer promotes low-rank structures across

different views, effectively preserving shared and complementary information among views (Gao et al. 2020). Similarly, we also introduce an adaptive weighting scheme to quantify the view contributions. Thus, compressed anchor graph learning for multi-view remote sensing data can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{U}^v, \mathbf{S}^v, \beta^v} \sum_{v=1}^V \frac{1}{\beta_v} \|\mathbf{U}^v - \mathbf{U}^v \mathbf{S}^v\|_F^2 + \lambda \|\mathcal{S}\|_{\mathbb{S}}^p \quad (4) \\ \text{s.t. } \mathbf{s}_i^{vT} \mathbf{1} = 1, s_{ij}^v \geq 0, \beta^T \mathbf{1} = 1, \beta_v > 0, \\ \text{rank}(\mathbf{L}_{S^v}) = M - C, \end{aligned}$$

where  $\mathbf{L}_{S^v} = \mathbf{D}_{S^v} - (\mathbf{S}^{vT} + \mathbf{S}^v)/2$  is the normalized Laplacian matrix related to  $\mathbf{S}^v$ ,  $\mathbf{D}_{S^v}$  is the degree matrix with diagonal elements  $(d_{S^v})_{ii} = \sum_j (s_{ij}^v + s_{ji}^v)/2$ , and  $\frac{1}{\alpha_v}$  denotes the view weight. The first four constraints are nonnegativity and normalization constraints for  $\mathbf{S}$  and  $\beta$ , while the final constraint is the connectivity constraint for  $\mathbf{S}^v$ .

**Theorem 1** *The multiplicity  $C$  of the eigenvalue 0 of the normalized Laplacian matrix  $\mathbf{L}_{S^v}$  equals the number of connected components in the graph associated with  $\mathbf{S}^v$ .*

According to Theorem 1, if  $\text{rank}(\mathbf{L}_{S^v}) = M - C$ ,  $\mathbf{S}^v$  has exactly  $C$  connected components, each corresponding to a distinct cluster. Consequently, the weighted Laplacian matrix also satisfies  $\text{rank}(\mathbf{L}_S) = M - C$ , where  $\mathbf{L}_S = \sum_{v=1}^V \frac{1}{\beta_v} \mathbf{L}_{S^v}$ . Under this condition, the implicitly weighted consensus graph  $\mathbf{S} = \sum_{v=1}^V \frac{1}{\beta^v} \mathbf{S}^v$  preserves the same property, i.e., it has exactly  $C$  connected components (Chen et al. 2022). Therefore, the  $C$ -cluster structure among anchors can be directly uncovered. The constraint  $\text{rank}(\mathbf{L}_{S^v}) = M - C$  is non-convex and thus difficult for direct optimization. Given that  $\mathbf{L}_{S^v}$  is positive semi-definite, all its eigenvalues satisfy  $\sigma_i(\mathbf{L}_{S^v}) \geq 0$ . Therefore, the rank constraint can be relaxed by minimizing the sum of its smallest  $C$  eigenvalues, i.e., enforcing  $\sum_{i=1}^C \sigma_i(\mathbf{L}_{S^v}) = 0$ . Based on Ky Fan's theorem (Fan 1949), we have  $\sum_{i=1}^C \sigma_i(\mathbf{L}_{S^v}) = \min_{\mathbf{F}^T \mathbf{F} = \mathbf{I}} \text{Tr}(\mathbf{F}^T \mathbf{L}_{S^v} \mathbf{F})$ , where  $\mathbf{F} \in \mathbb{R}^{M \times C}$  denotes the embedding matrix subject to an orthogonality constraint. Based on these observations, Eq. (4) can be reformulated as follows:

$$\begin{aligned} \min_{\mathbf{U}^v, \mathbf{S}^v, \beta^v} \sum_{v=1}^V \frac{1}{\beta_v} \|\mathbf{U}^v - \mathbf{U}^v \mathbf{S}^v\|_F^2 + \lambda \|\mathcal{S}\|_{\mathbb{S}}^p + 2\gamma \text{tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}), \quad (5) \\ \text{s.t. } \mathbf{s}_i^{vT} \mathbf{1} = 1, s_{ij}^v \geq 0, \beta^T \mathbf{1} = 1, \beta_v > 0, \mathbf{F}^T \mathbf{F} = \mathbf{I}. \end{aligned}$$

where  $\gamma$  denotes a regularization parameter.

**Overall Objective Function** The separation between anchor graph learning and compressed anchor graph learning may result in suboptimal performance. Thus, we combine Eqs. (3) and (5) into a unified framework to formulate the overall objective function of ETAP as follows:

$$\begin{aligned} \min_{\mathbf{W}^v, \mathbf{Z}^v, \mathbf{S}^v, \mathbf{U}^v, \beta^v, \alpha^v} \sum_{v=1}^V \frac{1}{\alpha_v} \sum_{i=1}^N \sum_{j=1}^M \left\| \mathbf{W}^{vT} \mathbf{x}_i^v - u_j^v \right\|_2^2 z_{ij}^v + \delta \sum_{v=1}^V \|\mathbf{Z}^v\|_F^2 \\ + \sum_{v=1}^V \frac{1}{\beta_v} \|\mathbf{U}^v - \mathbf{U}^v \mathbf{S}^v\|_F^2 + \lambda \|\mathcal{S}\|_{\mathbb{S}}^p + 2\gamma \text{tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) \\ \text{s.t. } \mathbf{z}_i^{vT} \mathbf{1} = 1, z_{ij}^v \geq 0, \mathbf{s}_i^{vT} \mathbf{1} = 1, s_{ij}^v \geq 0, \alpha^v \mathbf{1} = 1, \quad (6) \\ \alpha_v > 0, \beta^T \mathbf{1} = 1, \beta_v > 0, \mathbf{F}^T \mathbf{F} = \mathbf{I}, \mathbf{W}^{vT} \mathbf{S}_t^v \mathbf{W}^v = \mathbf{I}. \end{aligned}$$

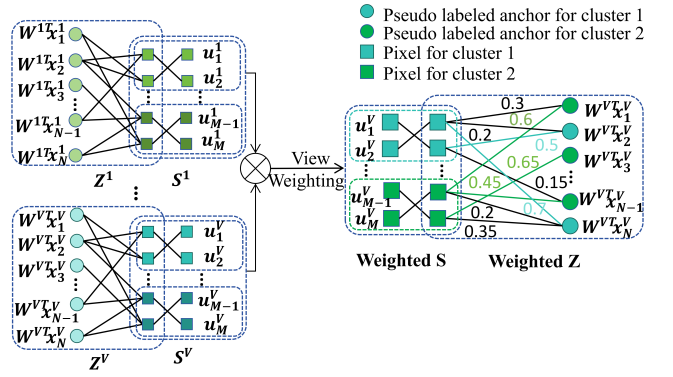


Figure 2: Illustration of the designed affinity propagation.

The joint learning of  $\{\mathbf{Z}^v\}_{v=1}^V$  and  $\{\mathbf{S}^v\}_{v=1}^V$  facilitates the discovery of their optimal structures. Through iterative optimization, the clustering structures of anchors gradually emerges as in the exactly  $C$  connected components in the consensus graph  $\mathbf{S} = \sum_{v=1}^V \frac{1}{\beta^v} \mathbf{S}^v$ .

**Affinity Propagation for Pixel Partitioning** Finally, the clustering labels of pixels can be revealed through affinity propagation based on the structured anchors in  $\mathbf{S}$  and the weighted anchor graph  $\mathbf{Z} = \sum_{v=1}^V \frac{1}{\alpha_v} \mathbf{Z}^v$ . Specifically, a multi-view pixel is assigned the same label as the anchor that exhibits the maximum affinity with the pixel in  $\mathbf{Z}$ . Thus, the clustering structures of pixels are efficiently obtained. Figure 2 illustrates the affinity propagation strategy used for pixel label inference.

## Optimization

Motivated by augmented Lagrange multiplier (ALM) (Lin, Liu, and Su 2011), we introduce an auxiliary variable  $\mathcal{J}$  and set  $\mathcal{J} = \mathcal{S}$ . Then, Eq. (6) can be rewritten as

$$\begin{aligned} \min_{\Omega} \sum_{v=1}^V \frac{1}{\alpha_v} \sum_{i=1}^N \sum_{j=1}^M \left\| \mathbf{W}^{vT} \mathbf{x}_i^v - u_j^v \right\|_2^2 z_{ij}^v + \delta \sum_{v=1}^V \|\mathbf{Z}^v\|_F^2 \\ + \sum_{v=1}^V \frac{1}{\beta_v} \|\mathbf{U}^v - \mathbf{U}^v \mathbf{S}^v\|_F^2 + \lambda \|\mathcal{J}\|_{\mathbb{S}}^p + 2\gamma \text{tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) \\ + \langle \mathcal{Y}, \mathcal{S} - \mathcal{J} \rangle + \frac{\mu}{2} \|\mathcal{S} - \mathcal{J}\|_F^2 \quad (7) \end{aligned}$$

$$\begin{aligned} \text{s.t. } \mathbf{z}_i^{vT} \mathbf{1} = 1, z_{ij}^v \geq 0, \mathbf{s}_i^{vT} \mathbf{1} = 1, s_{ij}^v \geq 0, \alpha^v \mathbf{1} = 1, \\ \alpha_v > 0, \beta^T \mathbf{1} = 1, \beta_v > 0, \mathbf{F}^T \mathbf{F} = \mathbf{I}, \mathbf{W}^{vT} \mathbf{S}_t^v \mathbf{W}^v = \mathbf{I}. \end{aligned}$$

where  $\Omega = \{\mathbf{W}^v, \mathbf{Z}^v, \mathbf{S}^v, \mathbf{U}^v, \beta^v, \alpha^v, \mathcal{J}\}$  are variables to be optimized,  $\mathcal{Y}(\cdot, v, \cdot) = \mathcal{Y}^{(v)}$  is the Lagrange multiplier, and  $\mu > 0$  is the penalty factor. Our model reliably converges by iteratively updating each variable individually while fixing the others during optimization.

**1) Updating  $\mathbf{Z}^v$ :** When  $\mathbf{W}^v, \mathbf{U}^v, \mathbf{S}^v, \mathbf{F}, \alpha^v, \beta^v$  and  $\mathcal{J}$  are fixed, the optimization for  $\mathbf{Z}^v$  becomes

$$\begin{aligned} \min_{\mathbf{Z}^v} \frac{1}{\alpha_v} \sum_{i=1}^N \sum_{j=1}^M \left\| \mathbf{W}^{vT} \mathbf{x}_i^v - u_j^v \right\|_2^2 z_{ij}^v + \delta \sum_{i=1}^N \sum_{j=1}^M z_{ij}^v \\ \text{s.t. } \mathbf{z}_i^{vT} \mathbf{1} = 1, z_{ij}^v \geq 0. \quad (8) \end{aligned}$$

---

**Algorithm 1: ETAP**


---

**Input:** Multi-view data  $\{\mathbf{X}^v\}_{v=1}^V$ , number of anchors  $M$ , cluster number  $C$ , projected dimension  $\{R_v\}_{v=1}^V$

**Output:** Clustering results.

- 1: Obtain  $\{\mathbf{X}^v\}_{v=1}^V$  by performing spatial denoising;
  - 2: Initialize  $\{\mathbf{A}^v\}_{v=1}^V$  from spatial anchor generation;
  - 3: Initialize  $\{\mathbf{Z}^v\}_{v=1}^V$ ; Initialize  $\alpha_v$  and  $\beta_v$  with  $\frac{1}{V}$ ;
  - 4: **repeat**
  - 5:   Update  $\{\mathbf{W}^v\}_{v=1}^V$  by solving Eq. (11);
  - 6:   Update  $\{\alpha_v\}_{v=1}^V$  using Eq.(13);
  - 7:   Update  $\{\mathbf{Z}^v\}_{v=1}^V$  by solving Eq. (9);
  - 8:   Update  $\{\mathbf{S}^v\}_{v=1}^V$  by solving Eq. (19);
  - 9:   Update  $\mathbf{F}$  by solving Eq. (16);
  - 10:   Update  $\mathcal{J}$  using Eq. (17);
  - 11:    $\mathcal{Y} = \mathcal{Y} + \mu(\mathcal{S} - \mathcal{J})$ ;
  - 12:    $\mu = \min(\eta\mu, \mu_{\max})$ ;
  - 13:   Update  $\{\beta_v\}_{v=1}^V$  using Eq. (15);
  - 14:   Update  $\{\mathbf{U}^v\}_{v=1}^V$  using Eq. (22);
  - 15: **until** Convergence
  - 16: **return**  $\{\mathbf{Z}^v\}_{v=1}^V$  and  $\{\mathbf{S}^v\}_{v=1}^V$ .
- 

Denoting  $b_{ij}^v = \frac{1}{\alpha_v} \|\mathbf{W}^{vT} \mathbf{x}_i^v - \mathbf{u}_j^v\|_2^2$ , Eq. (8) can be solved by independently updating  $z_i^v$  as follows:

$$\min_{z_i^v} \frac{1}{2} \left\| z_i^v + \frac{\mathbf{b}_i^v}{2\delta} \right\|_2^2, \quad \text{s.t. } z_i^v \mathbf{1} = 1, z_{ij}^v \geq 0. \quad (9)$$

There is a closed-form solution to problem (9) as proved in (Nie et al. 2016).

**2) Updating  $\mathbf{W}^v$ :** When  $\mathbf{Z}^v, \mathbf{U}^v, \mathbf{S}^v, \mathbf{F}, \alpha_v, \beta_v$  and  $\mathcal{J}$  are fixed, the optimization for  $\mathbf{W}^v$  becomes:

$$\min_{\mathbf{W}^v} \frac{1}{\alpha_v} \sum_{i=1}^N \sum_{j=1}^M \left\| \mathbf{W}^{vT} \mathbf{x}_i^v - \mathbf{W}^{vT} \mathbf{a}_j^v \right\|_2^2 z_{ij}^v + \frac{1}{\beta_v} \left\| \mathbf{W}^{vT} \mathbf{A}^v - \mathbf{W}^{vT} \mathbf{A}^v \mathbf{S}^v \right\|_F^2. \quad (10)$$

According to (Wang et al. 2022) and  $\mathbf{U}^v = \mathbf{W}^{vT} \mathbf{A}^v$ , problem (10) can be rewritten as:

$$\min_{\mathbf{W}^v} \text{tr} \left( \mathbf{W}^{vT} (\mathbf{P}^v + \mathbf{Q}^v) \mathbf{W}^v \right). \quad (11)$$

$\mathbf{P}^v = \frac{1}{\alpha_v} (\mathbf{X}^v \mathbf{D}_z^{\text{row}} \mathbf{X}^{vT} - 2\mathbf{X}^v \mathbf{Z} \mathbf{A}^{vT} + \mathbf{A}^v \mathbf{D}_z^{\text{col}} \mathbf{A}^{vT})$  with  $\mathbf{D}_z^{\text{row}} = \text{diag}(\sum_{j=1}^M z_{ij}^v)$  and  $\mathbf{D}_z^{\text{col}} = \text{diag}(\sum_{i=1}^N z_{ij}^v)$ , and  $\mathbf{Q}^v = \frac{1}{\beta_v} (\mathbf{A}^v \mathbf{A}^{vT} - 2\mathbf{A}^v \mathbf{S}^v \mathbf{A}^{vT} + \mathbf{A}^v \mathbf{S}^v \mathbf{S}^{vT} \mathbf{A}^{vT})$ . The optimal  $\mathbf{W}^v$  is formed by the eigenvectors of  $\mathbf{S}_t^{v-1} (\mathbf{P}^v + \mathbf{Q}^v)$  corresponding to  $R_v$  smallest eigenvalues.

**3) Updating  $\alpha^v$ :** When  $\mathbf{W}^v, \mathbf{U}^v, \mathbf{S}^v, \mathbf{F}, \mathbf{Z}^v, \beta_v$  and  $\mathcal{J}$  are fixed, the optimization for  $\alpha^v$  becomes:

$$\min_{\alpha^v} \frac{1}{\alpha_v} \sum_{i=1}^N \sum_{j=1}^M \left\| \mathbf{W}^{vT} \mathbf{x}_i^v - \mathbf{u}_j^v \right\|_2^2 z_{ij}^v. \quad (12)$$

Letting  $h_v = \sqrt{\sum_{i=1}^N \sum_{j=1}^M \left\| \mathbf{W}^{vT} \mathbf{x}_i^v - \mathbf{u}_j^v \right\|_2^2 z_{ij}^v}$ , Eq. (12) can be solved according to Cauchy-Schwartz's inequality (Nie, Wang, and Huang 2014) as follows:

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

**4) Updating  $\beta_v$ :** When  $\mathbf{W}^v, \mathbf{U}^v, \mathbf{S}^v, \mathbf{F}, \mathcal{J}, \alpha_v$  and  $\mathbf{Z}^v$  are fixed, the optimization for  $\beta_v$  becomes

$$\min_{\beta_v} \frac{1}{\beta_v} \|\mathbf{U}^v - \mathbf{U}^v \mathbf{S}^v\|_F^2, \quad \text{s.t. } \beta^T \mathbf{1} = 1, \beta_v > 0. \quad (14)$$

Similar to the solution for  $\alpha^v$ , we denote  $e^v = \|\mathbf{U}^v - \mathbf{U}^v \mathbf{S}^v\|_F$ , and update  $\beta^v$  as follows:

$$\beta_v = \frac{e^v}{\sum_{v=1}^V e^v}. \quad (15)$$

**5) Updating  $\mathbf{F}$ :** When  $\mathbf{W}^v, \mathbf{U}^v, \mathbf{S}^v, \mathbf{Z}^v, \alpha_v, \beta_v$  and  $\mathcal{J}$  are fixed, the optimization for  $\mathbf{F}$  becomes

$$\min \text{tr} (\mathbf{F}^T \mathbf{L}_S \mathbf{F}), \quad \text{s.t. } \mathbf{F}^T \mathbf{F} = \mathbf{I}. \quad (16)$$

The optimal solution of  $\mathbf{F}$  is composed of the eigenvectors of  $\mathbf{L}_S$  corresponding to  $C$  smallest eigenvalues.

**6) Updating  $\mathcal{J}$ :** When  $\mathbf{W}^v, \mathbf{U}^v, \mathbf{S}^v, \mathbf{Z}^v, \alpha_v, \beta_v$  and  $\mathbf{F}$  are fixed, the optimization for  $\mathcal{J}$  becomes

$$\begin{aligned} \mathcal{J}^* &= \min_{\mathcal{J}} \lambda \|\mathcal{J}\|_{\oplus}^p + \langle \mathcal{Y}, \mathcal{S} - \mathcal{J} \rangle + \frac{\mu}{2} \|\mathcal{S} - \mathcal{J}\|_F^2 \\ &= \min_{\mathcal{J}} \frac{\lambda}{\mu} \|\mathcal{J}\|_{\oplus}^p + \frac{1}{2} \left\| \mathcal{J} - \left( \mathcal{S} + \frac{\mathcal{Y}}{\mu} \right) \right\|_F^2. \end{aligned} \quad (17)$$

Problem (17) has a closed-form solution based on Theorem 3 from (Gao et al. 2021).

**7) Updating  $\mathbf{S}^v$ :** When  $\mathbf{W}^v, \mathbf{U}^v, \mathbf{Z}^v, \mathbf{F}, \alpha_v, \beta_v$  and  $\mathcal{J}$  are fixed, let  $\mathbf{Q}^v = \mathbf{J}^v - \frac{\mathbf{Y}^v}{\mu}$ . The optimization for  $\mathbf{S}^v$  becomes

$$\begin{aligned} \min_{\mathbf{S}^v} \frac{1}{\beta_v} \|\mathbf{U}^v - \mathbf{U}^v \mathbf{S}^v\|_F^2 + \frac{\gamma}{\beta_v} \sum_{i,j=1}^M \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 s_{ij}^v \\ + \frac{\mu}{2} \|\mathbf{S}^v - \mathbf{Q}^v\|_F^2 \\ \text{s.t. } \mathbf{s}_i^{vT} \mathbf{1} = 1, s_{ij}^v \geq 0. \end{aligned} \quad (18)$$

Expanding Eq. (18) into trace form, we have :

$$\begin{aligned} \min_{\mathbf{S}^v} \sum_{i=1}^M \mathbf{s}_i^{vT} \left( \frac{\mathbf{K}^v}{\beta_v} + \frac{\mu}{2} \mathbf{I} \right) \mathbf{s}_i^v + \left( \frac{-2}{\beta_v} \mathbf{k}_i^v + \frac{\gamma}{\beta_v} \mathbf{t}_i^T - \mu \mathbf{q}_i^{vT} \right) \mathbf{s}_i^v \\ \text{s.t. } \mathbf{s}_i^{vT} \mathbf{1} = 1, s_{ij}^v \geq 0. \end{aligned} \quad (19)$$

where  $t_{ij} = \|\mathbf{f}_i - \mathbf{f}_j\|_2^2$ , and  $\mathbf{K}^v = \mathbf{U}^{vT} \mathbf{U}^v$ . Many existing quadratic programming packages (Kang, Peng, and Cheng 2017) can be utilized to solve each  $\mathbf{s}_i^v$  of  $\mathbf{S}^v$  in problem (19).

**8) Updating  $\mathbf{U}^v$ :** When  $\mathbf{W}^v, \mathbf{S}^v, \mathbf{Z}^v, \mathbf{F}, \alpha_v, \beta_v$  and  $\mathcal{J}$  are fixed, the optimization for  $\mathbf{U}^v$  becomes

$$\min_{\mathbf{U}^v} \frac{1}{\alpha_v} \sum_{i=1}^N \sum_{j=1}^M \left\| \mathbf{W}^{vT} \mathbf{x}_i^v - \mathbf{u}_j^v \right\|_2^2 z_{ij}^v + \frac{1}{\beta_v} \|\mathbf{U}^v - \mathbf{U}^v \mathbf{S}^v\|_F^2. \quad (20)$$

Using the previously derived equations and removing irrelevant terms, we can derive the following equation:

$$\begin{aligned} \min_{\mathbf{U}^v} \text{Tr} \left[ \frac{1}{\alpha_v} \left( -2\mathbf{W}^{vT} \mathbf{X}^v \mathbf{Z}^v \mathbf{U}^{vT} + \mathbf{K}^v \mathbf{D}_z^{\text{col}} \right) + \right. \\ \left. \frac{1}{\beta_v} \left( \mathbf{K}^v - \mathbf{K}^v \mathbf{S}^v - \mathbf{K}^v \mathbf{S}^{vT} - \mathbf{S}^{vT} \mathbf{K}^v \mathbf{S}^v \right) \right]. \end{aligned} \quad (21)$$

The optimal solution of  $\mathbf{U}^v$  can be obtained by taking the derivative and setting it to zero. Thus, we have

$$\mathbf{U}^v = \beta_v \mathbf{W}^{vT} \mathbf{X}^v \mathbf{Z}^v \left[ \beta_v \mathbf{D}_z^{\text{col}} + \alpha_v (\mathbf{I} - \mathbf{S}^v - \mathbf{S}^{vT} + \mathbf{S}^v \mathbf{S}^{vT}) \right]^{-1}. \quad (22)$$

Finally, the entire optimization procedure for Eq. (6) is outlined in Algorithm 1.

Datasets	Metrics	MFLVC	DIVIDE	SCMVC	TMPCC	DFCPC	AMKSC	CDD	MSSAGF	SAMVGC	Ours
Berlin	ACC	0.4174	0.4315	0.4459	0.4214	0.5327	0.5298	<u>0.6447</u>	0.5988	0.4890	<b>0.7151</b>
	Kappa	0.2578	0.2806	0.2932	0.2342	0.3640	0.3798	<u>0.4286</u>	0.3587	0.3116	<b>0.5120</b>
	NMI	0.2629	0.3076	0.3068	0.3245	<u>0.3988</u>	0.3039	0.2841	0.2716	0.2888	<b>0.4404</b>
	Purity	0.6709	0.6969	0.6854	0.6843	<u>0.7087</u>	0.6645	0.6645	0.6008	0.7002	<b>0.7621</b>
	ARI	0.1525	0.2256	0.2192	0.2016	0.3298	0.2091	<u>0.3500</u>	0.3376	0.2501	<b>0.4994</b>
MUUFL	ACC	0.5118	0.5132	0.3796	0.5353	0.5610	0.4364	<u>0.5816</u>	0.4335	0.4897	<b>0.6659</b>
	Kappa	0.4215	0.4080	0.2911	0.4211	0.4357	0.3630	<u>0.4885</u>	0.1904	0.4115	<b>0.5531</b>
	NMI	0.3928	0.3480	0.3010	<b>0.5257</b>	0.4624	0.4236	0.4689	0.1944	0.3978	<u>0.5056</u>
	Purity	0.6361	0.5917	0.5660	0.6527	0.6285	0.6806	<u>0.7011</u>	0.4357	0.6769	<b>0.7295</b>
	ARI	0.2956	0.3573	0.1765	0.4158	<u>0.4368</u>	0.2694	0.4098	0.1368	0.2556	<b>0.6126</b>
MDAS	ACC	0.2997	0.2450	0.2565	0.2841	0.3248	0.2832	N/A	0.3181	<u>0.4215</u>	<b>0.5634</b>
	Kappa	0.2298	0.1874	0.1839	0.2100	0.1780	0.1871	N/A	0.0995	<u>0.2851</u>	<b>0.3039</b>
	NMI	0.2666	0.2332	0.2387	<u>0.2938</u>	0.2073	0.2359	N/A	0.1493	0.2832	<b>0.3092</b>
	Purity	0.6066	0.6059	0.6046	0.6022	0.5738	0.5960	N/A	0.4944	<b>0.6394</b>	<u>0.6239</u>
	ARI	0.1134	0.0773	0.0884	0.1077	0.1195	0.1192	N/A	0.0678	<u>0.1821</u>	<b>0.3008</b>

Table 1: Clustering performance on multi-view remote sensing datasets. Best results are in **bold**, second-best are underlined.

## Complexity Analysis

The optimization algorithm of our ETAP involves updating eight variables. To solve  $\{\mathbf{W}_v\}_{v=1}^V$ , it takes  $\mathcal{O}(\sum_{v=1}^V B_v R_v^2)$  to achieve eigenvalue decomposition and matrix inversion. To learn  $\mathbf{Z}^v$ , it needs  $\mathcal{O}(VNM)$  for calculation. The update of  $\mathbf{S}^v$  requires  $\mathcal{O}(VM^3)$  to get the closed-form solution. To update  $\{\mathbf{U}_v\}_{v=1}^V$ , it costs  $\mathcal{O}(RBNM^2)$ , where  $B = \sum_{v=1}^V B_v$  and  $R = \sum_{v=1}^V R_v$ . To solve  $\mathbf{F}$ , it requires  $\mathcal{O}(M^2C)$  to calculate the eigenvectors of  $\mathbf{L}_S$ . To update  $\mathcal{J}$ , it takes  $\mathcal{O}(2VM^2 \log(VM) + V^2M^2)$ . As for  $\alpha_v$  and  $\beta_v$ , their update trivially requires  $\mathcal{O}(1)$ . For clustering inference strategy, it takes  $\mathcal{O}(NMR)$  to get clustering results. Due to  $M \ll N$ , and  $V, C$  are small constants, the main computational complexity is  $\mathcal{O}(RBNM^2)$ . when  $R \ll N$  and  $B \ll N$  the proposed method shows a nearly linear time complexity of  $\mathcal{O}(N)$ .

## Experiments

### Experiment Setup

**Datasets.** Three datasets are utilized for experiments, including Berlin, MUUFL and MDAS. The Berlin dataset contains two views: an HS image with 244 spectral bands and a SAR image with 4 channels. It covers an area of  $1723 \times 476$  pixels and includes 8 land-cover categories. The MUUFL dataset comprises three views: an HS image with 180 spectral bands, a SAR image with 4 features, and a DSM image with 1 channel. The dataset size is  $332 \times 485$  pixels with 7 land-cover categories. The MDAS dataset is composed of four views: an HS image with 242 spectral bands, an MS image with 12 bands, a SAR image with 2 features, and an DSM image with 1 channel. The dataset spans  $300 \times 360$  pixels and includes 14 land-cover categories. These datasets span diverse resolutions, modalities and scenarios, offering a comprehensive testbed.

**Compared Methods.** To evaluate our ETAP, nine state-of-the-art methods are adopted for comparison, including MFLVC (Xu et al. 2022), DIVIDE (Lu et al. 2024), SCMVC (Yu et al. 2024), TMPCC (Cai et al. 2023), DFCPC (Xu et al. 2024), AMKSC (Cai et al. 2025), CDD (Liu and

Chang 2025), MSSAGF (Wang, Zhang, and Zhou 2025) and SAMVGC (Guan et al. 2025).

**Implementation Details.** The optimal parameters for ETAP is set by a grid-search strategy. Specifically,  $M$  is tuned within  $[150 : 25 : 300]$ , and  $p$  is varied from  $[0.1 : 0.1 : 1]$ .  $R_v$  is selected from  $[B_v/4 : 5 : B_v/2]$  for high-dimensional views and from  $[1 : 2 : B_v]$  for low-dimensional views. For  $\xi$ ,  $\gamma$  and  $\lambda$ , they are selected from  $[10^{-3}, 10^{-2}, \dots, 10^3]$ . For other baselines, we adopt their released code and recommended parameter settings. Clustering performance is evaluated using five common metrics, including accuracy (ACC), Kappa coefficient, normalized mutual information (NMI), purity and adjusted rand index (ARI). All experiments are carried out on a server configured with two NVIDIA RTX 4090 GPUs, each with 24GB of memory. Under optimal parameter settings, all reported results are averaged over ten independent runs to reduce randomness and ensure reliability.

### Comparison Results

**Clustering Performance.** As shown in Table 1, our ETAP achieves the optimal performance across all datasets in most cases. On the MUUFL dataset, it exceeds the second-best baseline CCD by 8.43%, 6.46%, 3.67%, 2.84% and 20.28% in terms of the five metrics. On the Berlin dataset, it surpasses CDD by 7.04%, 8.34%, 15.63%, 9.76% and 14.94%, demonstrating notable improvements over SAMVGC and DFCPC as well. On the MDAS dataset, CDD fails to handle more than two views, resulting in no derived results. Our ETAP outperforms the second-best method SAMVGC by 14.19%, 1.88%, 2.6% and 11.87% in terms of the four metrics, with the exception of Purity. These substantial and consistent gains demonstrate the superiority of our ETAP in leveraging multi-view information for improved clustering accuracy across diverse remote sensing scenarios.

To visualize the clustering effect, we display the clustering maps of different methods on the MUUFL datasets in Figure 3. Our ETAP aligns best with the ground truth. This is because the spatial structure alignment of anchors across views, enhancing the quality of anchor graph fusion.

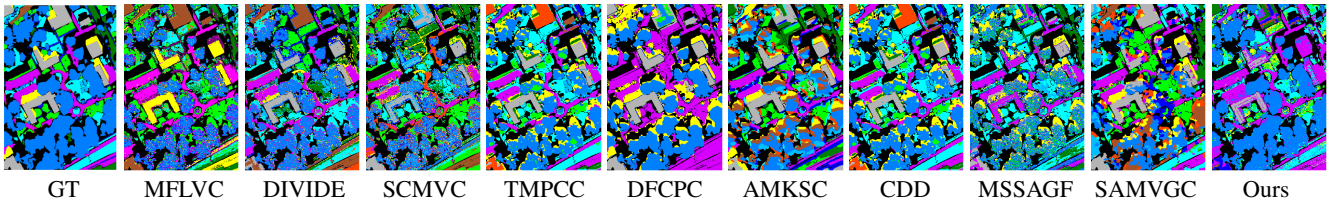


Figure 3: Clustering maps of the MUUFL dataset obtained by different methods. GT: Ground truth.

Methods	Berlin	MUUFL	MDAS
MFLVC	312.07	140.96	378.07
DIVIDE	188111.80	18696.60	3483.00
SCMVC	3081.72	1999.63	993.69
TMPCC	2238.89	891.45	477.50
DFCPC	8697.48	682.90	218.24
AMKSC	<b>149.59</b>	137.68	29.57
CDD	308.68	39.56	N/A
MSSAGF	1495.70	59.98	130.12
SAMVGC	239.82	179.01	1080.73
ETAP	214.96	<b>9.03</b>	<b>8.59</b>

Settings	S1	S2	S3	S4	Ours	
Factors	SG	✓	✓	✓	✓	✓
	SD	✗	✓	✓	✓	✓
	SP	✗	✗	✓	✓	✓
	PL	✗	✗	✗	✓	✓
	AMW	✗	✗	✗	✗	✓
Datasets	Berlin	0.5051	0.5317	0.5588	0.6241	<b>0.7151</b>
	MUUFL	0.3583	0.3786	0.4367	0.4527	<b>0.6659</b>
	MDAS	0.2721	0.3494	0.4565	0.4662	<b>0.5634</b>

Table 3: Ablation study of our ETAP (in ACC).

Table 2: Runtime record on multiple datasets (in seconds).

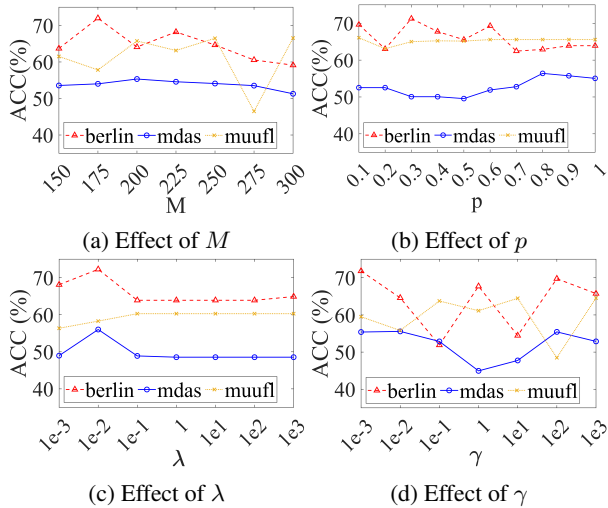


Figure 4: Performance variance of our ETAP.

**Running Time.** Table 2 reports the runtime comparison of different methods across three datasets. Our ETAP achieves the highest efficiency on the MUUFL and MDAS datasets, reducing execution time by 30.53 and 20.98 seconds compared to the second fastest methods. In addition, it obtains the second-best efficiency on the Berlin dataset. These results demonstrate the practical potential of our method for real-world applications.

## Model Analyses

**Parameter Sensitivity Study.** To investigate the parameter sensitivity of ETAP, we conduct experiments with different parameter settings and display the results in Figure 4. The results show that the superpixel size (i.e.,  $M$ ) should align

with dataset size, with values between 175 and 250 generally yielding good performance. The tensor Schatten  $p$ -norm offers stable results across a wide range of  $p$ . While  $\lambda$  has minimal impact,  $\gamma$  significantly affects performance and should be tuned for each dataset. Overall, proper parameter selection enhances the accuracy of ETAP.

**Ablation Study.** Table 3 presents the ablation study results. We use “✓” and “✗” to indicate whether the model includes or excludes each component. For clarity, SG denotes superpixel generation, SD represents spatial denoising, SP refers to the tensor Schatten  $p$ -norm constraint, PL stands for projection learning, and AMW indicates adaptive modality weighting. From Table 3, the model performance on all three datasets shows a consistent improvement as each component is gradually added. For example, on the Berlin dataset, “S2” surpasses “S1” by 2.66%, “S3” outperforms “S2” by 2.71%, “S4” improve upon S3 by over 6.53%, and our ETAP exceeds “S4” by 9.10%. Compared to the degraded settings, our ETAP with all components achieves the best performance across all three datasets, demonstrating the synergistic effect of each component in enhancing the model.

## Conclusion

In this paper, we proposed an efficient tensorized multi-view anchor graph clustering method with affinity propagation for remote sensing data. Based on superpixel-based denoising and anchor selection, anchor graph learning and compressed anchor graph learning with tensor regularization are jointly performed in a unified framework to uncover the clustering structures of anchors through a connectivity constraint. The designed affinity propagation strategy enables accurate and efficient partitioning for multi-view pixels. An alternating optimization algorithm is developed for the proposed formulation. Extensive experiments demonstrate the superiority of the proposed method.

## Acknowledgments

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

## References

- Cai, Y.; Zhang, Z.; Ghamisi, P.; Ding, Y.; Liu, X.; Cai, Z.; and Gloaguen, R. 2022. Superpixel Contracted Neighborhood Contrastive Subspace Clustering Network for Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–13.
- Cai, Y.; Zhang, Z.; Ghamisi, P.; Rasti, B.; Liu, X.; and Cai, Z. 2023. Transformer-Based Contrastive Prototypical Clustering for Multimodal Remote Sensing Data. *Information Sciences*, 649: 119655.
- Cai, Y.; Zhang, Z.; Liu, X.; Ding, Y.; Li, F.; and Tan, J. 2025. Learning Unified Anchor Graph for Joint Clustering of Hyperspectral and LiDAR Data. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4): 6341–6354.
- Chen, Z.; Zhang, C.; Mu, T.; and He, Y. 2022. Tensorial Multiview Subspace Clustering for Polarimetric Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–13.
- Fan, K. 1949. On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations I. *Proceedings of the National Academy of Sciences*, 35(11): 652–655.
- Gao, Q.; Wan, Z.; Liang, Y.; Wang, Q.; Liu, Y.; and Shao, L. 2020. Multi-View Projected Clustering with Graph Learning. *Neural Networks*, 126: 335–346.
- Gao, Q.; Zhang, P.; Xia, W.; Xie, D.; Gao, X.; and Tao, D. 2021. Enhanced Tensor RPCA and its Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6): 2133–2140.
- Guan, R.; Tu, W.; Wang, S.; Liu, J.; Hu, D.; Tang, C.; Feng, Y.; Li, J.; Xiao, B.; and Liu, X. 2025. Structure-Adaptive Multi-View Graph Clustering for Remote Sensing Data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 16933–16941.
- Kang, Z.; Peng, C.; and Cheng, Q. 2017. Twin Learning for Similarity and Clustering: A Unified Kernel Approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).
- Kilmer, M. E.; and Martin, C. D. 2011. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3): 641–658.
- Li, J.; Gao, Q.; Wang, Q.; and Xia, W. 2024a. Tensorized label learning on anchor graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 13537–13544.
- Li, X.; Pan, Y. P.; Sun, Y.; Sun, Q.; Sun, Y.; W. Tsang, I.; and Ren, Z. 2025. Incomplete Multi-view Clustering with Paired and Balanced Dynamic Anchor Learning. *IEEE Transactions on Multimedia*, 7087–7098.
- Li, X.; Pan, Y. P.; Sun, Y.; Sun, Q. S.; Tsang, I. W.; and Ren, Z. 2024b. Fast Unpaired Multi-view Clustering. *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 4488–4496.
- Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation. In *Advances in Neural Information Processing Systems 24*, 612–620.
- Liu, M.-Y.; Tuzel, O.; Ramalingam, S.; and Chellappa, R. 2011. Entropy Rate Superpixel Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2097–2104.
- Liu, S.; and Chang, L. 2025. Conditional Dual Diffusion for Multimodal Clustering of Optical and SAR Images. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(6): 5318–5330.
- Liu, S.; Liao, Q.; Wang, S.; Liu, X.; and Zhu, E. 2024. Robust and Consistent Anchor Graph Learning for Multi-View Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36(8): 4207–4219.
- Lu, P.; Jiang, X.; Zhang, Y.; Liu, X.; Cai, Z.; Jiang, J.; and Plaza, A. 2023. Spectral-Spatial and Superpixelwise Unsupervised Linear Discriminant Analysis for Feature Extraction and Classification of Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–15.
- Lu, Y.; Lin, Y.; Yang, M.; Peng, D.; Hu, P.; and Peng, X. 2024. Decoupled Contrastive Multi-View Clustering with High-Order Random Walks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 14193–14201.
- Luo, F.; Liu, Y.; Gong, X.; Nan, Z.; and Guo, T. 2024. EMVCC: Enhanced Multi-View Contrastive Clustering for Hyperspectral Images. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 6288–6296.
- Nie, F.; Wang, X.; and Huang, H. 2014. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 977–986. ACM.
- Nie, F.; Wang, X.; Jordan, M. I.; and Huang, H. 2016. The Constrained Laplacian Rank Algorithm for Graph-Based Clustering. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI)*, 1969–1976.
- Shi, G.; Huang, H.; and Wang, L. 2020. Unsupervised Dimensionality Reduction for Hyperspectral Imagery via Local Geometric Structure Feature Learning. *IEEE Geoscience and Remote Sensing Letters*, 17(8): 1425–1429.
- Sun, Y.; Li, X.; Sun, Q.; Zhang, M.; Ren, Z.; Cai, J.; Kankanhalli, M.; Prabhakaran, B.; Boll, S.; Subramanian, R.; Zheng, L.; Singh, V.; Cesar, P.; Xie, L.; and Xu, D. 2022. Structure Diversity-Induced Anchor Graph Fusion for Multi-View Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(12): 5682–5695.
- Wan, Y.; Ma, A.; Zhang, L.; and Zhong, Y. 2022. Multi-objective Sine Cosine Algorithm for Remote Sensing Image Spatial-Spectral Clustering. *IEEE Transactions on Cybernetics*, 52(10): 11172–11186.
- Wang, H.; and Zhu, S. 2020. Multisource Aggregation Search and Scheduling for Remote Sensing Data Cluster. *IEEE Geoscience and Remote Sensing Letters*, 17(2): 352–356.

- Wang, L.; Zhang, Y.; Li, H.; Liu, X.; and Chen, J. 2024. Adaptive Fusion of Multi-view Remote Sensing Data for Optimal Sub-field Crop Yield Prediction. *arXiv preprint arXiv:2401.11844*.
- Wang, R.; Nie, F.; Wang, Z.; He, F.; and Li, X. 2019. Scalable Graph-Based Clustering With Nonnegative Relaxation for Large Hyperspectral Image. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10): 7352–7364.
- Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2022. Fast Parameter-Free Multi-View Subspace Clustering With Consensus Anchor Guidance. *IEEE Transactions on Image Processing*, 31: 556–568.
- Wang, X.; Zhang, Y.; and Zhou, Y. 2025. Multi-modal Remote Sensing Image Clustering with Multi-scale Spectral-Spatial Anchor Graphs. *IEEE Transactions on Geoscience and Remote Sensing*, 63(4405612): 1–12.
- Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; and He, L. 2022. Multi-Level Feature Learning for Contrastive Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16051–16060.
- Xu, S.; Ding, X.; Zhang, Y.; Zhang, Z.; Gao, H.; and Zhang, B. 2024. Dual-Feature Attention-Based Contrastive Prototypical Clustering for Multimodal Remote Sensing Data. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–13.
- Xue, Z.; Tan, X.; Yu, X.; Liu, B.; Yu, A.; and Zhang, P. 2022. Deep Hierarchical Vision Transformer for Hyperspectral and LiDAR Data Classification. *IEEE Transactions on Image Processing*, 31: 3095–3110.
- Yang, C.; Li, Z.; and Zhang, L. 2023. Bootstrapping Interactive Image-Text Alignment for Remote Sensing Image Captioning. *arXiv:2312.01191*.
- Yang, X.; Cao, W.; Lu, Y.; and Zhou, Y. 2023. QTN: Quaternion Transformer Network for Hyperspectral Image Classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(12): 7370–7384.
- Yu, H.; Bian, H.-X.; Chong, Z.-L.; Liu, Z.; and Shi, J.-Y. 2024. Multi-View Clustering with Semantic Fusion and Contrastive Learning. *Neurocomputing*, 603: 128264.
- Zhang, L.; and Zhang, L. 2022. Artificial Intelligence for Remote Sensing Data Analysis: A Review of Challenges and Opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 10(2): 270–294.
- Zhang, M.; Li, W.; Du, Q.; Gao, L.; and Zhang, B. 2020. Feature Extraction for Classification of Hyperspectral and LiDAR Data Using Patch-to-Patch CNN. *IEEE Transactions on Cybernetics*, 50(1): 100–111.
- Zhang, M.; Li, W.; Zhang, Y.; Tao, R.; and Du, Q. 2023. Hyperspectral and LiDAR Data Classification Based on Structural Optimization Transmission. *IEEE Transactions on Cybernetics*, 53(5): 3153–3164.
- Zhang, Y.; Yan, S.; Jiang, X.; Zhang, L.; Cai, Z.; and Li, J. 2024a. Dual Graph Learning Affinity Propagation for Multimodal Remote Sensing Image Clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–13.
- Zhang, Y.; Yan, S.; Zhang, L.; and Du, B. 2024b. Fast Projected Fuzzy Clustering With Anchor Guidance for Multimodal Remote Sensing Imagery. *IEEE Transactions on Image Processing*, 33: 4640–4653.