

# Superpixel-Based and Spatially Regularized Diffusion Learning for Unsupervised Hyperspectral Image Clustering

Kangning Cui<sup>ID</sup>, Ruoning Li, Sam L. Polk, Yinyi Lin, Hongsheng Zhang<sup>ID</sup>, James M. Murphy<sup>ID</sup>, Robert J. Plemmons<sup>ID</sup>, and Raymond H. Chan<sup>ID</sup>

**Abstract**—Hyperspectral images (HSIs) provide exceptional spatial and spectral resolution of a scene, crucial for various remote sensing applications. However, the high dimensionality, presence of noise and outliers, and the need for precise labels of HSIs present significant challenges to the analysis of HSIs, motivating the development of performant HSI clustering algorithms. This article introduces a novel unsupervised HSI clustering algorithm—superpixel-based and spatially regularized diffusion learning (S<sup>2</sup>DL)—which addresses these challenges by incorporating rich spatial information encoded in HSIs into diffusion geometry-based clustering. S<sup>2</sup>DL employs the entropy rate superpixel (ERS) segmentation technique to partition an image into superpixels, then constructs a spatially regularized diffusion graph using the most representative high-density pixels. This approach reduces computational burden while preserving accuracy. Cluster modes, serving as exemplars for underlying cluster structure, are identified as the highest-density pixels farthest in diffusion distance from other highest-density pixels. These modes guide the labeling of the remaining representative pixels from ERS superpixels. Finally, majority voting is applied to the labels assigned within each superpixel to propagate labels to the rest of the image. This spatial-spectral approach simultaneously simplifies graph construction, reduces computational cost, and improves clustering performance. S<sup>2</sup>DL's performance is illustrated with extensive experiments on four publicly available, real-world HSIs: Indian Pines, Salinas, Salinas A, and WHU-Hi. Additionally, we apply S<sup>2</sup>DL to landscape-scale, unsupervised mangrove species mapping in the Mai Po

Nature Reserve (MPNR), Hong Kong, using a Gaofen-5 HSI. The success of S<sup>2</sup>DL in these diverse numerical experiments indicates its efficacy on a wide range of important unsupervised remote sensing analysis tasks.

**Index Terms**—Diffusion geometry, hyperspectral image (HSI) clustering, spatial regularization, species mapping, superpixel segmentation.

## I. INTRODUCTION

**H**YPERSPECTRAL images (HSIs) encode reflectance across a broad spectrum of wavelengths in the visual and infrared light spectra, storing a rich characterization of large spatial regions with high spectral resolution. These images can be obtained through various platforms, such as airplanes, drones, or orbital spectrometers [1], [2]. HSI data have been shown to be useful in a wide range of high-impact applications, e.g., the identification of land use and land cover [3], [4], [5], [6], [7], [8], the unmixing of spectral signatures [9], [10], [11], [12], and the fusion of images from different sources and modalities [13], [14], [15], [16], [17]. These applications often leverage machine learning and deep learning approaches to exploit the rich information stored within HSIs.

Machine learning and deep learning have significantly improved HSI classification with advanced models that handle the complex spatial-spectral aspects of HSIs [18], [19], [20], [21], [22]. Convolutional, graph convolutional, and recurrent neural networks have redefined accuracy and efficiency benchmarks in supervised HSI classification tasks [23], [24], [25], [26], [27], [28]. However, the need for specialized knowledge and extensive fieldwork for expert annotations makes HSI labeling a resource-intensive and costly task, motivating the development of unsupervised clustering algorithms for analyzing HSIs [1], [5], [29]. Nevertheless, even HSI clustering algorithms, which rely upon no expert annotations or ground-truth (GT) labels for image segmentations, face at least two key challenges in practice [30], [31], [32], [33].

The first challenge for HSI clustering is the sheer volume of pixels in an image that the algorithm must analyze. HSIs being large datasets are a challenge independent of their inherent high dimensionality, which stems from the high spectral resolution obtained using advanced hyperspectral spectrometers to generate high-quality HSI data. Large images are especially challenging for graph-based HSI clustering

Manuscript received 24 December 2023; revised 4 March 2024; accepted 2 April 2024. Date of publication 4 April 2024; date of current version 17 April 2024. This work was supported in part by HKRGC GRF under Grant CityU1101120, Grant CityU11309922, and Grant CRF C1013-21GF; in part by HKRGC-NSFC under Grant N-CityU214-19; in part by the National Science Foundation under Grant DMS-2309519 and Grant DMS-2318894; and in part by HKRGC under Grant HKU27602020 and Grant HKU17613022. (Corresponding author: Kangning Cui.)

Kangning Cui and Raymond H. Chan are with the Department of Mathematics, City University of Hong Kong, Hong Kong, and also with Hong Kong Centre for Cerebro-Cardiovascular Health Engineering, Hong Kong (e-mail: kangnicui2-c@my.cityu.edu.hk; raymond.chan@cityu.edu.hk).

Ruoning Li is with the Department of Mathematics, City University of Hong Kong, Hong Kong (e-mail: ruoningli3-c@my.cityu.edu.hk).

Sam L. Polk was with the Department of Mathematics, Tufts University, Medford, MA 02115 USA (e-mail: sam.polk@outlook.com).

Yinyi Lin and Hongsheng Zhang are with the Department of Geography, The University of Hong Kong, Hong Kong (e-mail: yinyilin@hku.hk; zhanghs@hku.hk).

James M. Murphy is with the Department of Mathematics, Tufts University, Medford, MA 02115 USA (e-mail: JM.Murphy@tufts.edu).

Robert J. Plemmons is with the Department of Computer Science, Wake Forest University, Winston-Salem, NC 27106 USA (e-mail: plemmons@wfu.edu). Digital Object Identifier 10.1109/TGRS.2024.3385202

methods, which construct a graph to represent pixel affinities and can scale quadratically, if not worse, with the number of pixels if the graph is not constructed carefully [1], [4], [34], [35]. Several techniques have been proposed to mitigate the computational demands of the graph construction process. For instance, anchor-based graph approaches reduce the number of graph nodes by selecting representative anchor points, while the Nyström extension approximates the graph's eigendecomposition to streamline later analysis [36], [37], [38]. However, the selection of anchors, often through random or  $K$ -Means methods, can misrepresent the dataset, neglect spatial information, and be noise-sensitive [39]. Addressing the computational challenges, it is important to recognize that HSIs typically possess a low intrinsic dimensionality because of the natural relationships between spectral bands, which can often be captured in a manifold coordinate system [40], [41], [42]. This reduced intrinsic dimensionality motivates the downsampling of pixels, as a carefully selected subset of the HSI is likely to contain sufficient information to approximate the low-dimensional geometry intrinsic to an HSI [40].

A second major challenge to HSI clustering algorithms is the presence of noise and outliers in HSIs, which often results in a significant reduction in performance on important unsupervised material classification problems due to internal and external factors. These factors include sensor noise, atmospheric effects, and spectral variability, resulting in a noisy characterization of a scene in which it can be difficult to differentiate materials in an unsupervised setting [43], [44]. A further complication is the high intra-class spectral variability often observed in different locations of an HSI resulting from variance in illumination conditions and viewing angles [9]. Integrating spatial information with spectral information can help mitigate the effects of spectral variability. This allows the algorithm to use contextual information from spatially nearby pixels, which often belong to the same class and share similar spectral characteristics [45], [46], [47].

In this article, we propose superpixel-based and spatially regularized diffusion learning (S<sup>2</sup>DL): a superpixel-based diffusion learning (DL) approach to unsupervised clustering of large HSIs. Our approach first employs entropy rate superpixel (ERS) segmentation to partition the image into spatial regions of similarly expressive pixels. ERS segmentation assumes local homogeneity and is meant to reduce the effect of spatial-spectral variability within each superpixel. S<sup>2</sup>DL uses a kernel density estimate (KDE) to identify a small number of most-representative pixels from each superpixel for use in the construction of a spatially regularized  $k$ -nearest neighbor (kNN) graph [41], [48], [49], [50]. Importantly, the graph used in S<sup>2</sup>DL explicitly incorporates spatial information into graph construction by allowing edges only between pixels and their nearest neighbors within a spatial radius. S<sup>2</sup>DL locates and assigns unique labels to single pixels from each cluster to serve as cluster exemplars: highest-density pixels farthest in spatially regularized diffusion distances [41], [48], [49], [50] from other high-density pixels. S<sup>2</sup>DL then propagates labels across the graph using a local backbone (LBB) spread and diffusion-based label propagation. After the clustering of pixels representative of superpixels, majority voting is performed

within each superpixel, ensuring spatial homogeneity of cluster assignments. As will become clear in extensive numerical experiments showing its efficacy, S<sup>2</sup>DL's procedure results in a substantial decrease in computational complexity while at the same time mitigating the effects of noise and outliers. Thus, S<sup>2</sup>DL is a highly efficient and accurate approach to unsupervised HSI clustering.

To summarize, our contributions are threefold. First, we present an efficient approach that merges superpixel segmentation with the selection of representative pixels, significantly reducing computational demands for graph-based analysis. Second, we introduce spatially regularized graphs that capitalize on the inherent spatial regularity of HSIs, thereby enhancing clustering efficacy. Lastly, we employ the LBB spread to further incorporate modal pixels into downstream non-modal labeling. This article is organized as follows. In Section II, we discuss related work on HSI clustering, diffusion geometry, and spatial-spectral HSI clustering algorithms while over-viewing terminology and background. Section III introduces S<sup>2</sup>DL and motivates its approach to spatial-spectral clustering (SC) of HSIs. Section IV shows the efficacy of S<sup>2</sup>DL through extensive numerical experiments comparing S<sup>2</sup>DL with classical and state-of-the-art unsupervised approaches on four benchmark HSI datasets: Indian Pines, Salinas, Salinas A, and WHU-Hi. Additionally, this section showcases the application of S<sup>2</sup>DL to a Gaofen-5 sensor-collected HSI dataset for unsupervised mapping mangrove species in Mai Po Nature Reserve (MPNR), Hong Kong [51]. Section V concludes and discusses future work.

## II. RELATED WORKS

### A. Overview of HSI Clustering Techniques

Clustering is an unsupervised learning technique that groups similar objects or data points without the need for GT labels or expert annotations [52]. We denote pixels in an HSI as a set  $X = \{x_i\}_{i=1}^N \subset \mathbb{R}^B$ , where each  $x_i$  denotes the spectral signature of a pixel in the image, and  $B$  represents the number of spectral bands. HSI clustering algorithms partition pixels into a clustering  $\{X_k\}_{k=1}^K$  (where each  $X_k$  is a cluster) such that pixels from the same cluster are “similar” (possibly due to a similar material constitution), while pixels from different clusters are “dissimilar” (containing different materials) [53]. The specific notion of similarity used varies widely across the many clustering algorithms in the literature.

Traditional clustering methods—e.g.,  $K$ -Means [54], Gaussian mixture model [55], and density-based spatial clustering of applications with noise [56]—have been employed for HSI clustering [4]. However, these methods often encounter challenges due to the presence of spectrally mixed pixels and noise inherent in HSI data [41], [43], [53], e.g., sensitivity to initialization, assumptions on the distribution of data [40], [41], [57], [58], and sensitivity to noise [4]. Density peak clustering (DPC) [59] was introduced to mitigate some of these distribution assumption errors (specifically, that of uniform density across clusters). DPC locates  $K$  points in a dataset to serve as cluster modes: exemplars for latent underlying cluster structure. DPC cluster modes are defined as

the highest-density points farthest in Euclidean distance from other high-density points. These cluster modes are assigned unique labels, which are propagated across the dataset by assigning (in order of decreasing density) each point the label of its  $\ell^2$ -nearest neighbor of higher density that is already labeled. However, the use of Euclidean distances to make pairwise comparisons between HSI pixels in DPC [59] and other traditional clustering algorithms [52] has been shown to reduce clustering quality on datasets with nonlinear decision boundaries between latent clusters [40], [49], [53], or due to the “curse of dimensionality,” that all pixels tend to appear roughly equidistant from one another in high-dimensional space [4], [41]. Finally, these methods often cluster individual pixels, agnostic to rich spatial information present in HSI data [1], [4], [9], [45], [49].

Deep clustering algorithms have been investigated for use on HSIs due to their ability to capture nonlinear decision boundaries and learn discriminative features for material classification in recent years [4], [5], [16], [60], [61], [62], [63], [64]. The vast deep clustering techniques in the literature range from contrastive learning-based [61], [62] to graph-based approaches [63], [64], [65]. The integration of superpixels in deep HSI analysis has been shown to enhance computational efficiency and semantic consistency [62], [66]. Despite their success on a wide range of HSI clustering problems, deep clustering algorithms are often highly sensitive to noise and perturbations in the data [67], [68], and recent research has indicated that their success may be attributable to preprocessing steps rather than the learning capabilities of the network itself [69]. Moreover, despite their highly accurate recovery of latent material structure in some applications, many deep HSI clustering approaches face limitations such as high computational complexity of training and the need for large training sets [60], [62], [64].

### B. DL for HSI

DL has emerged as a highly effective approach to extracting and utilizing the inherent geometric structure contained within HSI data in an unsupervised clustering framework [41], [42], [49], [53], [58]. DL interprets HSI pixels as nodes in an undirected, weighted graph, the edges between which indicate pairwise similarity between pixels [41], [70]. This graph can be stored in a sparse adjacency matrix  $\mathbf{W} \in \mathbb{R}^{N \times N}$ , where  $\mathbf{W}_{ij} = 1$  if the pixel  $x_j$  is one of the  $k_n$   $\ell^2$ -nearest neighbors (where  $\ell^2$  denotes the Euclidean distance) of the pixel  $x_i$  in  $X$ , and  $\mathbf{W}_{ij} = 0$  otherwise. DL relies on the data-dependent diffusion distance metric to make pairwise comparisons between pixels in the HSI [41], [42], enabling highly accurate extraction of latent nonlinear structure in HSIs [6], [41], [48], [53], [57], [71].

Diffusion distances are calculated by considering a Markov diffusion process on the graph underlying  $\mathbf{W}$  [42]. The transition matrix for this diffusion process can be calculated directly from  $\mathbf{W}$ :  $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ , where  $\mathbf{D}$  is the diagonal degree matrix with  $\mathbf{D}_{ii} = \sum_{j=1}^N \mathbf{W}_{ij}$ . Provided the graph underlying  $\mathbf{P}$  is irreducible and aperiodic,  $\mathbf{P}$  will possess a unique stationary distribution  $\pi \in \mathbb{R}^{1 \times N}$  such that  $\pi\mathbf{P} = \pi$ . The diffusion

distance at a time  $t \geq 0$  between any two pixels in the HSI  $x_i, x_j \in X$  [42], [48], [57], [58] is defined by

$$D_t(x_i, x_j) = \sqrt{\sum_{k=1}^N [(\mathbf{P}^t)_{ik} - (\mathbf{P}^t)_{jk}]^2 / \pi_k}. \quad (1)$$

Diffusion distances have a natural relationship to the clustering problem [42], [53], [58]. Indeed, one may expect many high-weight length- $t$  paths between pixels sampled from the same latent cluster but very few such paths between pixels sampled from different clusters, resulting in intra-cluster diffusion distances that are small relative to inter-cluster diffusion distances [42], [53], [58]. The diffusion time parameter  $t$  governs the scale of structure considered by diffusion distances, with smaller  $t$  enabling retrieval of small-scale local structure in the image and larger  $t$  retrieving global structure [42], [48], [57].

Importantly, the eigendecomposition of  $\mathbf{P}$  can be used for the efficient computation of diffusion distances. Indeed, given the eigenvalue–eigenvector pairs  $\{(\lambda_k, \psi_k)\}_{k=1}^N$  of the transition matrix  $\mathbf{P}$ , it can be shown [42] that

$$D_t(x_i, x_j) = \sqrt{\sum_{k=1}^N |\lambda_k|^{2t} [(\psi_k)_i - (\psi_k)_j]^2} \quad (2)$$

for  $t \geq 0$  and  $x_i, x_j \in X$ . Crucially, under the assumptions of irreducibility and aperiodicity,  $|\lambda_k| < 1$  for  $k > 1$ . This implies that, for sufficiently large  $t$ , diffusion distances can be accurately approximated using only the few eigenvectors  $\psi_k$  corresponding to the largest  $|\lambda_k|$  [42], [57], [58].

Employing diffusion geometry for HSI clustering yields significant advantages. Diffusion distances effectively counter the “curse of dimensionality” inherent to high-dimensional datasets like HSIs by extracting an intrinsic lower-dimensional, nonlinear representation of pixels, simultaneously reducing sensitivity to noise and redundant information [40], [41], [58]. Consequently, these methods provide a robust approach to extracting latent geometric structure hidden in high-dimensional HSI data [41], [48], [53] and have strong performance guarantees on clustering recovery across wide classes of data types [57], [58].

Furthermore, several studies have utilized random walks on graphs for HSI classification. These studies estimate the probabilities of transitioning from unlabeled to labeled pixels by minimizing energy derived from random walk processes, thereby facilitating the classification of unlabeled pixels [72], [73], [74]. However, DL distinguishes itself from these related algorithms through its temporal parameter, enabling the exploitation of both local and global data structures in HSIs effectively [42], [57]. Moreover, DL leverages eigendecomposition to efficiently estimate diffusion distances, thereby enhancing scalability [42], [58]. In addition, as will be discussed soon, there exist spatial–spectral variants of DL, achieved by constructing a spatially regularized graph or assigning labels via spatial window constraints, enhancing these algorithms’ sensitivity to the spatial context of each HSI pixel and therefore improving the quality of derived clusters [48], [49], [50], [75].



### C. Spatial–Spectral Analysis of HSI

Incorporating the spatial structure of an HSI into an HSI clustering algorithm is essential for achieving high-quality partitions and mitigating the effects of noise and spectral variability in HSIs. This is due to the tendency of nearby pixels to exhibit similar spectral properties [4], [37], [39], [60], [62], [63], [75]. Spatially regularized graphs offer a robust framework for embedding spatial context into graph-based HSI clustering algorithms like DL [48], [49]. Traditional graph-based methods that focus solely on spectral information and are agnostic to the spectral consistency observed within localized spatial regions in HSIs tend to perform poorly due to the heterogeneous spectral signatures and noise across an HSI scene [36], [39], [76]. Spatially regularized graphs counteract this spectral variance by limiting connections between pixels to those within a spatial radius around each pixel. Mathematically, a spatially regularized kNN graph may be defined through its corresponding weight matrix  $\mathbf{W}$ , with  $\mathbf{W}_{ij} = 1$  if  $x_j$  is one of the  $k_n$  Euclidean distance nearest neighbors of  $x_i$  from points within a  $(2R + 1) \times (2R + 1)$  spatial square centered at  $x_i$  in the HSI, where  $R \in \mathbb{N}$  is a user-defined spatial radius, and  $\mathbf{W}_{ij} = 0$  otherwise [48], [49], [53], [76]. Thus, spatially regularized graphs efficiently encode spatial coherence within the HSI by restricting edges within the graph underlying  $\mathbf{W}$  to spatially close pixels [4], [48], [49], [53], [76].

Spatially regularized graphs have emerged as a pivotal tool in enhancing HSI clustering and semi-supervised classification tasks [48], [49], [50]. The spatially regularized graph, specifically tailored for HSIs with diffusion distances and enhanced by a spatial window-based labeling consensus mechanism, was introduced in [49] to effectively leverage spatial context. Extending this approach, a multiscale framework was introduced, utilizing spatially regularized graphs to learn a single clustering scale that is most explanatory of latent multiscale cluster structure extracted by varying  $t$  in diffusion distances, as measured by the variation of information [48], [57]. Furthermore, spatially regularized graphs have found utility in active learning, a branch of semi-supervised learning that requires human input. Spatially regularized graphs aid in the strategic selection of a set of pixels for labeling—based on available budget—ensuring that the chosen pixels are locally coherent due to spatial constraints and globally representative through density estimation [50].

A second important field of research meant to incorporate spatial information into HSI clustering algorithms is that of superpixel segmentation. Superpixel segmentation algorithms partition the HSI into relatively small spatial regions exhibiting comparable spectral signatures, possibly due to similar mixtures of materials in the region they correspond to. These regions (or superpixels) capture local spatial structure, and the analysis of them (rather than that of the full HSI) reduces the number of pixels being analyzed, and hence, computational complexity associated with later analysis [77], [78]. Superpixel algorithms fall into two main classes: clustering-based, which iteratively cluster pixels based on convergence criteria [79], [80], [81], [82], and graph-based, which form superpixels by optimizing a cost function on a pixel-node graph [77], [78], [83], [84], [85]. Simple linear

iterative clustering (SLIC), as one of the most widely used clustering-based superpixel segmentation algorithm, adapts  $K$ -Means clustering to generate high-quality superpixels while maintaining a remarkably low computational and memory cost [79]. ERS is a representative graph-based segmentation algorithm that generates compact, homogeneous superpixels with similar sizes by minimizing the cost of cuts on the image graph, see Section III-B [77]. Given ERS's precision and enhanced boundary delineation, as demonstrated in superpixel segmentation benchmarks [86], we have chosen to integrate ERS into our proposed method.

Superpixel-based methods have been employed in HSI classification to incorporate spatial information into their predictions and reduce the effect of spatial–spectral variability within each superpixel [87], [88], [89], [90]. Several methods have combined superpixel segmentation with dimensionality reduction to address the high dimensionality of HSIs and simplify classification by using contextual pixel information. For instance, SuperPCA utilizes ERS and PCA at the superpixel level for enhanced local feature extraction from homogeneous areas, improving classification [46]. Similarly, superpixel-wise collaborative-representation graph embedding employs a Laplacian-regularized collaborative representation within superpixels for dimensionality reduction, tackling spectral redundancy and class variability [91]. Superpixel constraints aid in subspace learning and classification accuracy [92], and ERS effectively reduces noise and identifies anchor points for clustering [39]. Additionally, superpixel pooling autoencoders have been developed to capture superpixel-level latent representations to assist subsequent cluster analysis [62]. Despite these developments, the integration and exploration of superpixels in unsupervised HSI clustering are still relatively limited, particularly in conjunction with diffusion geometry [48], [49], [66], [75], [93].

## III. SUPERPIXEL-BASED AND SPATIALLY REGULARIZED DL

This section introduces the proposed algorithm for unsupervised HSI clustering:  $S^2DL$ . While pixel-wise clustering methods can be effective in certain scenarios, they often fail when applied to HSIs due to their disregard for spatial information: a key characteristic in HSI data. Ignoring the spatial context can lead to inaccurate clustering results as the spatial proximity of pixels often correlates with similarity in material composition in remotely sensed images.

### A. Overview of $S^2DL$

$S^2DL$  addresses this challenge by integrating spatial information into its clustering process in two key steps. First,  $S^2DL$  employs the graph-based ERS superpixel segmentation technique to obtain a high-quality superpixel segmentation of the HSI [77], [86], as described in Section III-B. Second, as described in Section III-C,  $S^2DL$  constructs a spatially regularized graph from carefully selected pixels from each ERS-derived superpixel, effectively capturing intrinsic spectral–spatial relationships between pixels within the HSI [41], [48], [49] while reducing overall computational

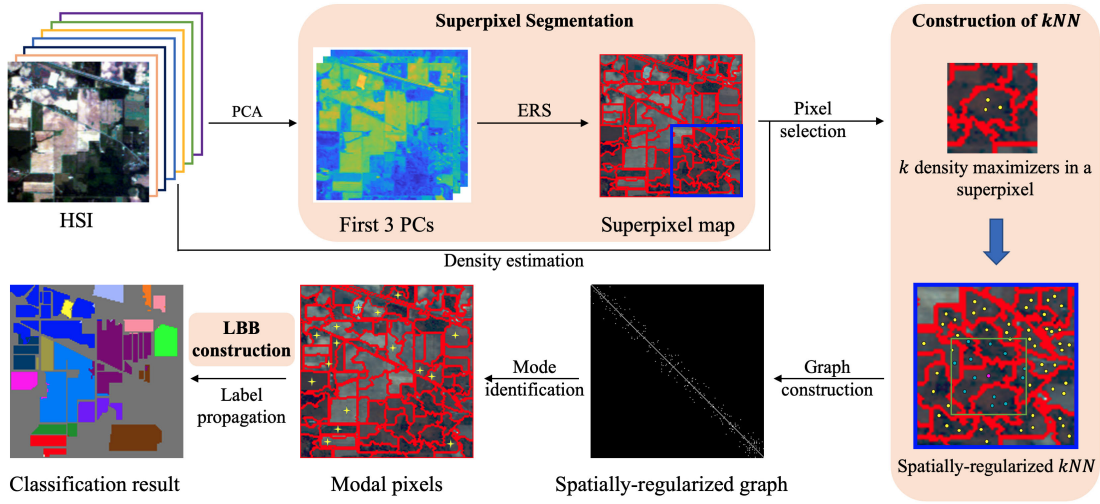


Fig. 1. Workflow of the S<sup>2</sup>DL algorithm. The algorithm begins by projecting the first three PCs of the input HSI, which are then used to create a superpixel map via the ERS algorithm. S<sup>2</sup>DL estimates the  $k$  highest-density pixels within each superpixel as representatives in graph construction, and then constructs a spatially regularized kNN graph. Mode pixels are subsequently identified and assigned unique labels, with the LBB of each mode receiving the same label as its respective mode. Then the labels are propagated to unlabeled selected pixels. The process concludes with majority voting within each superpixel to finalize the clustering.

complexity of graph construction and analysis. Finally, S<sup>2</sup>DL performs a diffusion-based clustering procedure to label superpixel exemplars and propagate those labels to the remaining image as described in Section III-D.

The combination of superpixel segmentation and the use of a spatially regularized graph is expected to result in superior clustering results using S<sup>2</sup>DL (as will be shown in our numerical experiments in Section IV). As is shown in the complexity analysis in Section III-E, S<sup>2</sup>DL's use of superpixel segmentation and spatially regularized diffusion distances offers a computationally efficient approach to HSI cluster analysis that is expected to translate to high-quality unsupervised analysis of large-scale HSI datasets. The detailed steps of S<sup>2</sup>DL are introduced in Sections III-B–III-E and summarized in Fig. 1 and Algorithm 2.

### B. ERS-Based Superpixel Segmentation

S<sup>2</sup>DL begins by computing a superpixel segmentation map using ERS, see Algorithm 1 [77]. As will become clear soon, the superpixels derived using ERS not only reflect spectral similarities between pixels in a graph-based procedure, but also the spatial information. Incorporating spatial information through superpixels not only offers a significant improvement over traditional pixel-wise clustering techniques, which often disregard spatial information, but also substantially reduces computational complexity by decreasing the number of pixels processed in subsequent graph construction.

ERS builds its superpixel segmentation based on an undirected weighted graph based on the projection of pixel spectra onto their first three principal components (PCs), denoted  $X_{\text{PCA}}$ . Mathematically, this initial graph may be defined using the weight matrix  $\mathbf{Z} \in \mathbb{R}^{N \times N}$  as

$$\mathbf{Z}_{ij} = \begin{cases} \exp\left(-\frac{m_{ij}}{2\sigma^2}\right), & \text{if } x_i \in \mathcal{N}_\ell(x_j) \text{ or } x_j \in \mathcal{N}_\ell(x_i) \\ 0, & \text{otherwise} \end{cases}$$

where  $m_{ij} = \|l(x_i) - l(x_j)\|_2^2 \cdot \|x_i - x_j\|_2^2$ , with  $l(x_i)$  and  $l(x_j)$  representing spatial coordinates of pixels  $x_i$  and  $x_j$ , respectively, and  $\sigma > 0$  is a tuning parameter meant to control interactions between pixels, and  $\mathcal{N}_\ell(x)$  is the set of  $\ell$  spatial nearest neighbors of  $x$  in the PCA-reduced feature space. In our later experiments, we set  $\sigma = 5$  and  $\ell = 8$ , the default values for these parameters [77]. Denote the edge set underlying this graph as  $E = \{(i, j) | \mathbf{Z}_{ij} > 0\}$ . ERS performs superpixel segmentation by locating a subset of edges  $A \subseteq E$  to form  $N_s$  compact, homogeneous, and well-balanced superpixels using the following optimization:

$$\begin{aligned} A^* = \underset{A}{\operatorname{argmax}} \quad & \mathcal{J}(A) = \mathcal{H}(A) + \alpha \mathcal{B}(A) \\ \text{s.t. } & A \subseteq E \text{ and } N_A \geq N_s \end{aligned} \quad (3)$$

where  $N_A$  is the number of connected components in  $A$  and  $\alpha > 0$  is the balancing factor that determines the trade-off between the terms  $\mathcal{H}(A)$  and  $\mathcal{B}(A)$  that are described below.

The term  $\mathcal{H}(A) = -\sum_{i=1}^N \mu_i \sum_{j=1}^N q_{ij}(A) \log(q_{ij}(A))$  measures the entropy rate of the edge set  $A$ , encouraging compact and homogeneous spatial regions in the ERS superpixel segmentation by considering a random walk on the graph subset  $A$  [77]. The entropy rate calculation relies on two main quantities: the  $N \times 1$  stationary distribution  $\mu \in \mathbb{R}^{N \times 1}$  of a random walk on the initial graph underlying  $\mathbf{Z}$ —defined by  $\mu_i = (\sum_{j=1}^N \mathbf{Z}_{ij} / \sum_{i,j=1}^N \mathbf{Z}_{ij})$ —and transition probabilities  $q_{ij}(A)$  on the edge subset  $A$  of  $E$ , defined as

$$q_{ij}(A) = \begin{cases} 1 - \frac{\sum_{k \in A_i} \mathbf{Z}_{ik}}{\sum_{k=1}^N \mathbf{Z}_{ik}}, & \text{if } i = j \\ \frac{\mathbf{Z}_{ij}}{\sum_{k=1}^N \mathbf{Z}_{ik}}, & \text{if } i \neq j \text{ and } (i, j) \in A \\ 0, & \text{if } i \neq j \text{ and } (i, j) \notin A \end{cases}$$

where  $A_i = \{k \in \{1, 2, \dots, N\} | (i, k) \in A\}$  denotes the set of indices of pixels directly connected to  $x_i$  in  $A$ . Thus, the entropy rate  $\mathcal{H}(A)$  increases monotonically with the addition

**Algorithm 1** ERS Algorithm [77]

---

**Input:**  $X_{PCA}$  (Projection of Pixels onto first 3 PCs),  
 $N_s$  (# superpixels),  $\alpha$  (# balancing factor)  
**Output:**  $S$  (Superpixel Map)

- 1 Construct the graph  $G = (X_{PCA}, E, \mathbf{Z})$ ;
- 2 Initialize the subset edge  $A^* = \emptyset$ ,  $U = E$  and  
 $S = \{S_i \mid S_i = \{x_i\}, 1 \leq i \leq N\}$ ;
- 3 **while**  $U \neq \emptyset$  **and**  $N_A \geq N_s$  **do**
- 4     Find edge  $\hat{a} = \operatorname{argmax}_a \mathcal{J}(A^* \cup \{a\}) - \mathcal{J}(A^*)$ ,  
       where  $\mathcal{J}(A^*) = \mathcal{H}(A^*) + \alpha \mathcal{B}(A^*)$
- 5     **if**  $A^* \cup \{\hat{a}\}$  is cycle-free **then**
- 6         Update  $A^* = A^* \cup \{\hat{a}\}$
- 7     **end**
- 8     Update  $U = U - \{\hat{a}\}$
- 9 **end**
- 10 **while**  $A \neq \emptyset$  **do**
- 11     **if**  $(i, j) \in A$  **then**
- 12         Find  $x_i \in S_a$  and  $x_j \in S_b$ , update  $S_a = S_a \cup S_b$
- 13         Update  $A = A - \{(i, j)\}$  and  $S = S - S_b$
- 14     **end**
- 15 **end**

---

of edges to  $A$ , particularly when these edges contribute to more coherent clusters [77].

The second term in ERS's optimization  $\mathcal{B}(A)$  is meant to encourage uniform superpixel size and a balanced number of superpixels. Mathematically,  $\mathcal{B}(A) = -\sum_{i=1}^{N_A} \mathbf{r}_i \log(\mathbf{r}_i) - N_A$ , where  $\mathbf{r} \in [0, 1]^{N_A}$  denotes the distribution of pixels in  $N_A$  connected components, i.e.,  $\mathbf{r}_i$  is the fraction of pixels in  $X_{PCA}$  that exist in connected component  $i$ . Therefore, all else equal, the incorporation of  $\mathcal{B}(A)$  downweights both complicated superpixel segmentations with high variation in superpixel size across the image, or segmentations with extraneous superpixels. Hence, the incorporation of  $\mathcal{B}(A)$  can be interpreted as balancing the resulting ERS superpixel segmentation, ensuring that new superpixels are roughly equal in scale when edges are added to  $A$  in the ERS optimization [77].

Together, these two terms in  $\mathcal{J}(A)$  monotonically increase as edges are added to  $A$ , referencing uniformly sized, spectrally similar spatial regions as derived superpixels. The proposed objective function is efficiently addressed using a "lazy greedy" heuristic that initializes with an empty edge set  $A$  and gradually adds edges to maximize the ERS objective function [94]. This iterative addition persists until the number of connected components  $N_A$  matches the pre-set target  $N_s$ , thereby achieving the specified number of superpixels [77]. Finally, superpixels are generated from the connected edges in the optimized edge set  $A^*$ . Mathematically, a superpixel  $S_i$  is a set of pixels with a coherent subset of edges in  $A^*$ , see details in Algorithm 1. In the proposed S<sup>2</sup>DL algorithm, the superpixel segmentation  $S$  is utilized for subsequent graph construction, as will become clear soon in Section III-C.

### C. Reduced Spatially Regularized Graph Construction

S<sup>2</sup>DL computes a spatially regularized graph (see Section II-C for details) using a small subset of carefully

selected pixels from the ERS superpixel segmentation. To find the representative pixels from each superpixel, S<sup>2</sup>DL first relies on the following quantity:

$$\zeta(x) = \frac{1}{\mathcal{Z}} \sum_{y \in k_n(x)} \exp(-\|x - y\|_2^2 / \sigma_0^2) \quad (4)$$

for each pixel  $x \in X$ , where  $k_n(x)$  denotes the set of  $k_n$   $\ell^2$ -nearest neighbors of the HSI pixel  $x$  in  $X$ ,  $\sigma_0 > 0$  is a scaling factor controlling the interaction radius between pixels, and the quantity  $\mathcal{Z}$  normalizes  $\zeta(x)$  to ensure that  $\sum_{y \in X} \zeta(y) = 1$ . Thus,  $\zeta(x)$  will be higher for modal pixels that are close in Euclidean distance to their  $k_n$   $\ell^2$ -nearest neighbors and small otherwise [41], [53], [58]. For each ERS superpixel, S<sup>2</sup>DL selects the  $k$  pixels within the superpixel maximizing  $\zeta(x)$ , resulting in  $N_s \times k$  representative pixels used for graph construction. Mathematically, we define this highly explanatory subset  $X_s$  of  $X$  by  $X_s = \bigcup_{i=1}^{N_s} \{x \in S_i \mid x \text{ is one of the } k \text{ maximizers of } \zeta(x) \text{ in } S_i\}$ .

Importantly, constructing the graph using  $X_s$  instead of the entire set  $X$  leads to a significant reduction in computational complexity, while maintaining an accurate characterization of essential spatial-spectral and geometric information. Once representative pixels have been identified, a sparse spatially regularized kNN adjacency graph is constructed from the pixels in  $X_s$ , following the procedure described in Section II-C. This graph encodes rich spatial information in an HSI data by restricting edges between pixels to spatial nearest neighbors [48], [50].

### D. Diffusion-Based Clustering

This section describes S<sup>2</sup>DL's unsupervised diffusion-based clustering procedure to obtain high-quality labeling of superpixels. S<sup>2</sup>DL is inspired by the ubiquitous DPC algorithm [59] (See Section II-A) and its diffusion geometry-based extensions [41], [53], [58] that rely on diffusion distances rather than Euclidean distances to perform cluster analysis. Specifically, S<sup>2</sup>DL uses diffusion distances calculated from the spatially regularized graph introduced in Section III-C for the unsupervised identification of cluster modes [48], [49], [58]. First, S<sup>2</sup>DL locates  $K$  pixels meant to serve as cluster modes—exemplars for latent cluster structure—and assigns these pixels unique labels. These cluster modes are identified as the  $K$  pixels in  $X_s$  that maximize  $\Delta_t(x) = d_t(x)\zeta(x)$ , where

$$d_t(x) = \begin{cases} \max_{y \in X_s} D_t(x, y), & x = \operatorname{argmax}_{y \in X_s} \zeta(y) \\ \min_{y \in X_s} \{D_t(x, y) \mid \zeta(y) \geq \zeta(x)\}, & \text{otherwise.} \end{cases} \quad (5)$$

In particular,  $d_t(x)$  is the diffusion distance at time  $t$  between  $x$  and that HSI pixel's  $D_t$ -nearest neighbor with higher density. Thus, cluster modes identified using S<sup>2</sup>DL are the highest-density pixels in  $X_s$  farthest in diffusion distance from other highest-density pixels [6], [41], [48], [49], [50], [53], [57], [58]. S<sup>2</sup>DL assigns  $\hat{C}(x_{m_k}) = k$  for  $k = 1, 2, \dots, K$ , where  $\{x_{m_k}\}_{k=1}^K$  are the  $K$  maximizers of



$\Delta_t(x)$  and  $\hat{C} \in \{1, 2, \dots, K\}^N$  is a clustering map with  $\hat{C}_i = k$  indicating that pixel  $x_i$  is assigned to cluster  $k$ .

After locating cluster modes, S<sup>2</sup>DL relies on an LBB to propagate modal labels to unlabeled pixels in  $X_s$  [95]. The LBB is pivotal, consisting not only of the cluster center but also of the nearest neighbors most likely to belong to the same cluster, as inferred through similarity measures such as Euclidean distance or cosine similarity. In S<sup>2</sup>DL, the LBB of each modal pixel—defined as the first  $k_n$  spatially regularized nearest neighbors of that cluster mode—is assigned the same label as that of the modal pixel [95]. The formation of the LBB plays a crucial role in the S<sup>2</sup>DL algorithm, highlighting the spatial coherence of HSI data early in the non-modal labeling process. Next, in order of descending density, the remaining pixels  $x \in X_s$  are labeled according to their  $D_t$ -nearest neighbor of higher density that is already labeled [49], [58], [59]:  $\hat{C}(x) = \hat{C}(x^*)$ , where

$$x^* = \underset{y \in X_s}{\operatorname{argmin}} \{D_t(x, y) | \hat{C}(y) > 0 \wedge \zeta(y) \geq \zeta(x)\}. \quad (6)$$

This label propagation is entirely unsupervised, requiring no input labels. Once all pixels in  $X_s$  have been labeled, S<sup>2</sup>DL propagates labels within to the rest of the image using a majority voting process, assigning the majority label of representative pixels in each superpixel to all pixels in a superpixel. Notably, the majority voting procedure further enforces the retention of essential spatial-spectral characteristics of the original HSI data in S<sup>2</sup>DL.

### E. Computational Complexity

This section analyzes S<sup>2</sup>DL (Algorithm 2) concerning its computational complexity and scaling. The first stage, wherein superpixels are calculated using ERS, has the following main components: computing the first three PCs at a complexity of  $\mathcal{O}(NB^2)$ , and the ERS algorithm itself. While the worst case complexity of ERS is  $\mathcal{O}(N^2 \log(N))$ , in practice, it often performs more efficiently, typically exhibiting an average-case complexity closer to  $\mathcal{O}(N \log(N))$  [77].

For nearest neighbor searches, we assume the use of cover trees, which enables efficient nearest neighbor searches in high-dimensional spaces [96]. Indeed, the computational complexity of searching for the  $k$   $\ell^2$ -nearest neighbors in  $X$  using cover trees is  $\mathcal{O}(k_n B c^d N \log(N))$ , where  $d$  is the HSI's doubling dimension [53], [96] and  $c > 0$  is a constant that is  $\mathcal{O}(1)$  with respect to the other parameters  $k_n$ ,  $B$ ,  $d$ , and  $N$ . In this complexity analysis, we assume that these two values remain constant across subsets of  $X$ , e.g., that the doubling dimension of  $X$  is the same as that of  $X_s$  [96]. Under this assumption, the computational complexity of computing the KDE at each pixel is  $\mathcal{O}(k_n B c^d N \log(N))$ . Similarly, locating the  $k$  KDE-maximizers from each superpixel has complexity  $\mathcal{O}(N_s \log(N/N_s) + \mathcal{O}(k N_s))$ . Finally, the computational complexity of building our reduced spatially regularized graph is  $\mathcal{O}(k_n B c^d k N_s \log(k N_s))$ , a notable reduction in computational complexity due to our earlier downsampling procedure.

To perform its diffusion-based clustering, S<sup>2</sup>DL requires  $\mathcal{O}(k_n B c^d k N_s \log(k N_s) + k_n L^2 k N_s)$  to calculate  $d_t(x)$ , where  $L$  is the number of eigenvectors of  $\mathbf{P}$  used to approximate

---

### Algorithm 2 S<sup>2</sup>DL Method

---

**Input:**  $X$  (HSI),  $N_s$  (# superpixels),  $k$  (# representative pixels per superpixel),  $\sigma_0$  (kernel scaling factor),  $k_n$  (# nearest neighbors),  $R$  (spatial radius),  $K$  (# of clusters)

**Output:**  $C$  (Clustering Map)

#### ERS-Based Superpixel Segmentation;

- 1 Calculate  $X_s$ : the projection of pixels in  $X$  onto its first three PCs;
- 2 Run ERS to segment the PCA-reduced HSI  $X_{PCA}$  into  $N_s$  superpixels;

#### Reduced Spatially Regularized Graph Construction;

- 3 Compute kernel density estimation  $\zeta(x) = \sum_{y \in k_n(x)} \exp(-||x - y||_2^2 / \sigma_0^2)$  for all  $x \in X$ ;
- 4 For each superpixel, store the  $k$  pixels in that superpixel maximizing  $\zeta(x)$  in  $X_s$ ;
- 5 Construct a spatially-regularized kNN adjacency graph using the selected  $k \cdot N_s$  pixels with spatial radius  $R$ ;

#### Diffusion-Based Clustering;

- 6 Compute  $\Delta_t(x) = \zeta(x) d_t(x)$  for  $x \in X_s$ , where  $d_t(x)$  is as in Equation 5;
  - 7 Identify the  $K$  maximizers of  $\Delta_t(x)$  as modal pixels and assign unique labels from 1 to  $K$ ;
  - 8 For each cluster mode, assign its LBB (with  $k_n$  nearest neighbors) the modal pixel's label;
  - 9 In order of decreasing density, assign each unlabeled pixel in  $X_s$  the label of their  $D_t$ -nearest neighbor of higher density  $\zeta$  that is already labeled (Equation 6);
  - 10 For each superpixel, assign all pixels in the superpixel the modal label among the  $k$  representative pixels in  $X_s$  from this superpixel;
- 

diffusion distances [42], [58]. In contrast to diffusion-based methods clustering all the pixels, S<sup>2</sup>DL therefore operates at a significantly reduced computational complexity through its analysis of superpixel exemplars. Then, the computational complexity of labeling the LBB and remaining pixels is  $\mathcal{O}(K k_n)$  and  $\mathcal{O}(k_n L B c^d k N_s \log(k N_s))$ , respectively [58], [95]. Finally, it costs  $\mathcal{O}(N_s k)$  to perform majority voting. Prior work has demonstrated that it is sufficient to take  $k_n = \mathcal{O}(\log(N))$  [53], [57], [58] and we expect that, for HSIs taken over the same scene,  $N_s = \mathcal{O}(1)$  and  $K = \mathcal{O}(1)$  with respect to  $N$ . Assuming that  $L = \mathcal{O}(1)$  and  $k = \mathcal{O}(\log(N))$  with respect to  $N$  also, S<sup>2</sup>DL's overall computational complexity reduces to  $\mathcal{O}(B c^d N \log^2(N))$ : log-linear with respect to the number of pixels. Notably, S<sup>2</sup>DL's computational complexity is dominated by its calculation of the KDE (not graph construction, as in other diffusion-based algorithms), indicating high-quality scaling to large-scale HSI clustering problems.

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section contains extensive numerical experiments showing the efficacy of the proposed S<sup>2</sup>DL algorithm.

We compared S<sup>2</sup>DL against both traditional and state-of-the-art comparison methods on four real-world HSIs often used for benchmarking new algorithms (Section IV-A). In addition, we demonstrate that S<sup>2</sup>DL may be applied to the real-world problem of landscape-scale species mapping of mangroves using remotely sensed HSI data collected over the MPNR (Section IV-B).

Among the algorithms we analyzed, several serve as baselines, as they do not integrate spatial information. *K*-Means [54] partitions data by minimizing intra-cluster  $\ell^2$ -distances. SC [70] employs *K*-Means on the first *K* eigenvectors of **P** to identify clusters. DPC [59] defines cluster modes as high-density points distant from others, assigning them unique labels that are then propagated to neighboring points based on decreasing density. PGDPC [97] discriminates many “peak” pixels with highest density among nearest neighbors as measured with a KDE from “non-peak” pixels, which are associated with their nearest neighbor of higher density. It then derives graph-based cluster assignments for peak pixels (relying on both density and pairwise geodesic distances for graph construction) and propagates along geodesic paths to remaining non-peak pixels [97]. DL [58]—recognized as an early diffusion-based clustering technique—relies on the metric  $\Delta_r(x)$  from Section III-D for identifying cluster modes, and performs label propagation by assigning to each point the label of its  $D_r$ -nearest neighbor with higher density. Diffusion and volume maximization-based image clustering (D-VIC) [53] is a recent diffusion-based clustering algorithm that incorporates spectral unmixing into a DL clustering framework, downweighting high-density, low-purity pixels in mode selection and non-modal labeling, while following DL for other aspects. Although shown to be successful on a wide range of HSI data [41], [53], [57], DL and D-VIC do not natively incorporate spatial information into their labeling procedures.

Other comparison algorithms implemented exploit both spatial and spectral information in their labeling procedures. Improved SC with Multiplicative Update Algorithm (SC-I) [37] modifies SC by iteratively solving the eigenvalue decomposition of the Laplacian matrix **L**, relaxing its discreteness condition, and integrating spatial context into the graph underlying **P**. SLIC-PGDPC (S-PGDPC) [97] extends PGDPC by taking the average of the spectral signatures within each superpixel and using this average as input for the PGDPC algorithm. Spectral-spatial diffusion learning (DLSS) [41] enhances DL by adopting a two-stage labeling scheme: initially, it assigns labels via DL, subject to a spatial consensus check through majority voting within a specified window, leaving non-conforming pixels unlabeled; this is followed by a secondary DL process to label the remaining unlabeled pixels. Spatial-spectral image reconstruction and clustering with diffusion geometry (DSIRC) builds on the D-VIC framework, utilizing both purity and density metrics for cluster mode identification, and enhances clustering accuracy by integrating spatial information through a shape-adaptive reconstruction process that effectively reduces noise before applying DL [75]. Spatially regularized DL (SRDL) [48], [49] further refines DLSS by constructing a spatially regularized graph, resulting

in clusters that are more spatially consistent. Hyperparameters for all algorithms were optimized for a grid search, as is described in Appendix.

To evaluate the performance of our S<sup>2</sup>DL algorithm, we employ a suite of metrics. Overall accuracy (OA) calculates the total fraction of pixels correctly clustered. Average accuracy (AA) measures the average OA across different classes. Cohen’s kappa coefficient ( $\kappa$ ), defined as  $\kappa = (p_o - p_e)/(1 - p_e)$ , contrasts observed accuracy against expected random accuracy [98]. Additionally, we introduce the sum of OA, AA, and  $\kappa$  (sum) as a composite metric to assess overall performance. We track the runtime (RT) in seconds to assess computational efficiency. The Hungarian algorithm is applied to align the clusters generated by S<sup>2</sup>DL and related algorithms with the GT labels, ensuring each cluster is accurately matched with its corresponding GT class for performance evaluation [99], and we determine the optimal clustering outcome by maximizing the sum. For all datasets, we set *K* as the GT number of clusters. All experiments were conducted in MATLAB R2021a with the same environment: Intel<sup>1</sup> Core<sup>2</sup> i7-10875H CPU @ 2.30 GHz, 8 cores, 64 GB RAM, run on a Windows 64-bit system. The code to replicate numerical experiments can be found at: <https://github.com/ckn3/S2DLgithub.com/ckn3/S2DL>.

#### A. Experiments on Benchmark HSI Datasets

1) *Benchmark Datasets*: In this section, we introduce the four benchmark datasets chosen for this study. These datasets, captured using the AVIRIS and Hyperspec sensors, serve as representatives of diverse agricultural landscapes and have been widely utilized for evaluating machine learning methods for HSI.

a) *Salinas and Salinas A*: Captured by the AVIRIS sensor in 1998, the spatially regular dataset Salinas showcases the agricultural terrains of Salinas Valley, California. It has a spectral range of 380 to 2500 nm across 224 bands, with spatial size  $512 \times 217$  pixels, totaling 111 104 pixels. The Salinas A subset zooms in on a specific region of the Salinas scene with  $83 \times 86$  pixels, totaling 7138 pixels. Gaussian noise was added for distinctiveness. While the broader Salinas dataset has 16 classes, Salinas A focuses on six main crop types.

b) *Indian Pines*: Produced by the AVIRIS sensor in 1992, this dataset portrays northwest Indiana farmlands. It covers a spectral range of 400 to 2500 nm over 224 bands and spreads over  $145 \times 145$  pixels, making up 21 025 pixels. Notably, it includes 16 GT classes, capturing various crops and infrastructure.

c) *WHU-Hi*: Collected in 2018 using the Headwall Nano-Hyperspec imaging sensor mounted on unmanned aerial vehicles, the dataset offers a detailed view of agricultural fields in LongKou, Hubei Province, China. It encompasses 270 spectral bands ranging from 400 to 1000 nm and covers a spatial extent of  $550 \times 400$  pixels, summing up to 220 000 pixels. The dataset is categorized into nine GT classes, which include six types of crops and three non-crop categories.

<sup>1</sup>Registered trademark.

<sup>2</sup>Trademarked.



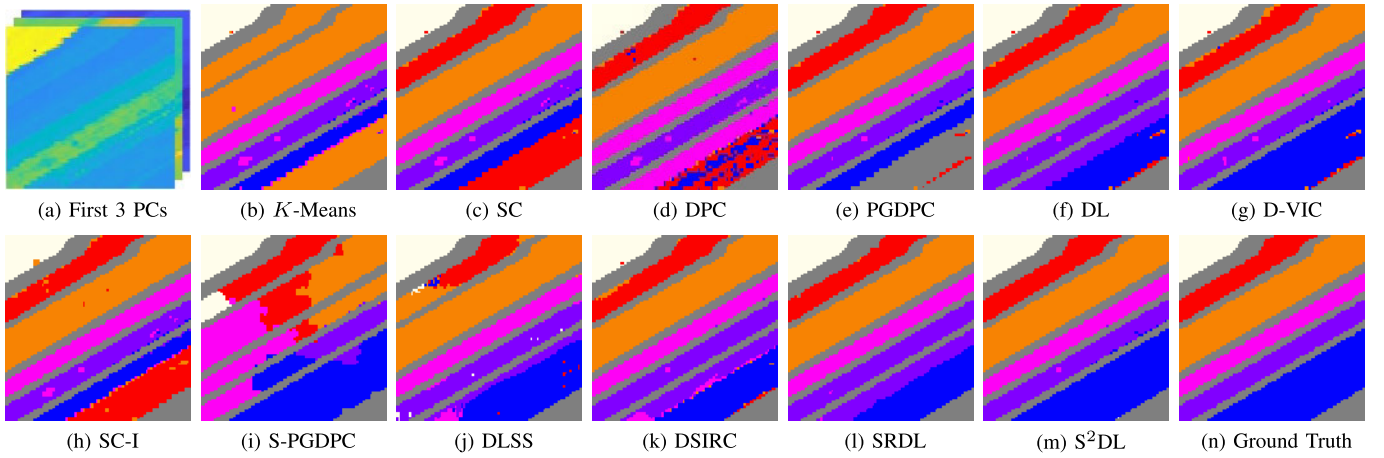


Fig. 2. Comparison of clustering results of (b)–(g) algorithms using spectral information and (h)–(l) algorithms using both spatial and spectral information and (m) S<sup>2</sup>DL on (a) Salinas A dataset with (n) GT.

2) *Numerical Results on Benchmark Datasets*: This section offers detailed comparisons between the clusterings produced by S<sup>2</sup>DL and various classical and state-of-the-art algorithms introduced earlier in this section. Across the four benchmark HSIs analyzed, S<sup>2</sup>DL delivers clusterings with the highest performance in terms of three evaluation metrics. Notably, it achieves almost perfect clustering on the Salinas A dataset, despite being unsupervised. Similarly, although the Indian Pines dataset is widely considered challenging due to its many classes being distributed widely across the scene, S<sup>2</sup>DL still manages to surpass its competitors in performance. On the Salinas dataset, S<sup>2</sup>DL not only yields the best performance in OA and  $\kappa$  but also completes the task in 8.80 s of RT. In contrast, the next best algorithm (SRDL) required 445.31 s of RT. This impressive improvement on RT is attributed to the use of superpixel-based reduction in graph size, illustrating S<sup>2</sup>DL's balanced approach to both speed and accuracy. For the WHU-Hi dataset, S<sup>2</sup>DL achieves the highest sum of OA, AA, and  $\kappa$ , exceeding the second best by 5.6%, and completes its analysis in only 15.46 s. Though the OA and  $\kappa$  are lower than the highest values, the AA surpasses the second best by 13.5%. These findings underscore S<sup>2</sup>DL's efficacy and efficiency in HSI clustering, marking it as a suitable choice for practical applications.

As visualized in Fig. 2, S<sup>2</sup>DL demonstrates notable precision in accurately identifying the GT labels within the Salinas A dataset. Whereas various algorithms split the cluster associated with the eight-week maturity romaine into two parts (visualized in dark blue in Fig. 2), both S<sup>2</sup>DL and the next highest-performing algorithm, D-VIC, correctly group these pixels into a unified cluster, with S<sup>2</sup>DL surpassing D-VIC by approximately 2% across all three metrics (see Table I). The precision demonstrated by D-VIC primarily stems from its incorporation of spectral unmixing information [53], whereas S<sup>2</sup>DL's effectiveness is attributed to its incorporation of both spatial and spectral data into its diffusion-based clustering procedure. Notably, S<sup>2</sup>DL mitigates spatial noise in its clustering through its superpixelation step, yielding a more spatially regularized clustering compared to D-VIC. This demonstrates the capability of S<sup>2</sup>DL to utilize spatial information as a robust alternative to the spectral unmixing in D-VIC, especially with spatially regular HSIs.

The performance of S<sup>2</sup>DL on the Salinas dataset is particularly noteworthy, having achieved the highest OA and  $\kappa$  scores in Table I. The AA of S<sup>2</sup>DL is marginally lower than that of SRDL (<3%), potentially because of SRDL's efficiency in recognizing smaller classes. Nevertheless, this slightly lower AA is compensated with over 5% higher OA and  $\kappa$ . Additionally, S<sup>2</sup>DL boasts a much lower RT due to its superpixelation step, further highlighting its efficiency. The gap in performance among S<sup>2</sup>DL, SRDL, and other methods indicates the significance of a spatially regularized graph approach in handling large and regular HSIs like Salinas.

As shown in Fig. 3, S<sup>2</sup>DL exhibits outstanding performance on the Indian Pines dataset as well. Indeed, S<sup>2</sup>DL achieved OA and  $\kappa$  values slightly higher than those of SRDL—its nearest competitor—by approximately 1% and AA values 4% in Table I. This improvement indicates S<sup>2</sup>DL's enhanced accuracy in classifying diverse classes, including those with fewer labels. Furthermore, S<sup>2</sup>DL delivers its efficient clustering performance with substantially lower RT than SRDL. As discussed earlier, this significant reduction in computational cost is due to S<sup>2</sup>DL's utilization of superpixelization. While SRDL utilizes a spatially regularized graph, S<sup>2</sup>DL integrates both superpixels and a spatially regularized graph. This combination not only facilitates more efficient use of spatial information but also optimizes the computational process, leading to the observed RT advantage.

Finally, S<sup>2</sup>DL distinguishes itself on the WHU-Hi dataset by achieving the best sum of metrics and the highest AA by 5.6% and 13.5% compared to DPC, its nearest competitor. DL and DPC, despite outperforming S<sup>2</sup>DL in OA and  $\kappa$  slightly, require extensive processing times due to pixel-wise analysis, exceeding 1800 s. Conversely, S<sup>2</sup>DL leverages superpixels for a swift analysis, concluding in merely 15.46 s. S<sup>2</sup>DL's lower OA and  $\kappa$  on WHU-Hi result from the limited utility of spatial regularization for this dataset, a contrast to its benefits on other datasets. This aspect is detailed later in Section IV-A4.

3) *Hyperparameter Robustness Analysis*: This section considers the robustness of S<sup>2</sup>DL's clustering performance to the selection of hyperparameters. We focus on parameters such as the number of superpixels  $N_s$ , the spatial radius  $R$ , the number of representative pixels per superpixel  $k$ , and the

TABLE I

COMPARISON OF UNSUPERVISED CLUSTERING METHODS ON BENCHMARK HSI DATASETS. THE BEST AND SECOND-BEST PERFORMANCES ARE INDICATED BY BOLD AND UNDERLINED VALUES, RESPECTIVELY. S<sup>2</sup>DL STANDS OUT FOR ITS HIGH-QUALITY CLUSTERING ACROSS ALMOST ALL METRICS AND DATASETS, WITH THE EXCEPTION OF AA ON SALINAS, AND OA &  $\kappa$  ON WHU. THE RELATIVELY LOW RTs AND TOP SUM SCORES COMPARED TO OTHER ALGORITHMS ALSO HIGHLIGHT THE SCALABILITY OF S<sup>2</sup>DL AND EFFICACY FOR ANALYSIS OF LARGE HSIs

Dataset		Method											
		<i>K</i> -Means	SC	DPC	PGDPC	DL	D-VIC	SC-I	S-PGDPC	DLSS	DSIRC	SRDL	S <sup>2</sup> DL
Salinas A	OA	0.764	0.841	0.786	0.844	0.887	<u>0.976</u>	0.827	0.647	0.890	0.911	0.895	<b>0.996</b>
	AA	0.749	0.887	0.849	0.893	0.920	<u>0.973</u>	0.875	0.680	0.888	0.903	0.926	<b>0.996</b>
	$\kappa$	0.703	0.806	0.740	0.813	0.860	<u>0.970</u>	0.789	0.568	0.862	0.889	0.870	<b>0.995</b>
	Sum	2.216	2.534	2.375	2.550	2.667	<u>2.919</u>	2.491	1.895	2.640	2.703	2.691	<b>2.987</b>
	RT	0.05	1.59	2.66	1.63	1.93	4.89	6.43	0.10	5.27	26.39	14.99	1.78
Indian Pines	OA	0.386	0.382	0.391	0.428	0.404	0.471	0.496	0.477	0.467	0.620	<u>0.640</u>	<b>0.647</b>
	AA	0.398	0.368	0.376	0.399	0.401	0.376	0.304	0.530	0.462	0.549	<u>0.553</u>	<b>0.591</b>
	$\kappa$	0.315	0.313	0.304	0.351	0.313	0.383	0.394	0.431	0.400	0.573	<u>0.596</u>	<b>0.602</b>
	Sum	1.099	1.063	1.071	1.178	1.118	1.230	1.194	1.438	1.329	1.742	<u>1.789</u>	<b>1.840</b>
	RT	1.14	14.40	13.10	19.75	13.64	24.33	70.43	0.38	20.55	136.02	30.52	2.19
Salinas	OA	0.639	0.662	0.668	–	0.687	0.696	–	0.590	0.702	0.677	<u>0.834</u>	<b>0.887</b>
	AA	0.612	0.633	0.654	–	0.662	0.623	–	0.487	0.674	0.612	<b>0.756</b>	<u>0.729</u>
	$\kappa$	0.597	0.620	0.627	–	0.646	0.653	–	0.551	0.662	0.633	<u>0.813</u>	<b>0.874</b>
	Sum	1.848	1.915	1.949	–	1.995	1.972	–	1.628	2.038	1.922	<u>2.403</u>	<b>2.490</b>
	RT	4.81	414.44	432.98	–	450.82	496.37	–	1.20	504.88	3059.74	445.31	8.80
WHU	OA	0.625	0.743	<b>0.857</b>	–	<u>0.857</u>	0.779	–	0.364	0.837	0.829	0.771	0.822
	AA	0.487	0.507	0.540	–	<u>0.540</u>	0.468	–	0.515	0.523	0.415	0.480	<b>0.675</b>
	$\kappa$	0.545	0.674	<b>0.810</b>	–	<u>0.810</u>	0.710	–	0.276	0.784	0.764	0.698	0.766
	Sum	1.657	1.924	<b>2.207</b>	–	<u>2.207</u>	1.957	–	1.155	2.144	2.008	1.948	<b>2.263</b>
	RT	13.96	1896.55	1881.26	–	1851.27	1965.99	–	2.50	2059.11	9755.97	2881.06	15.46

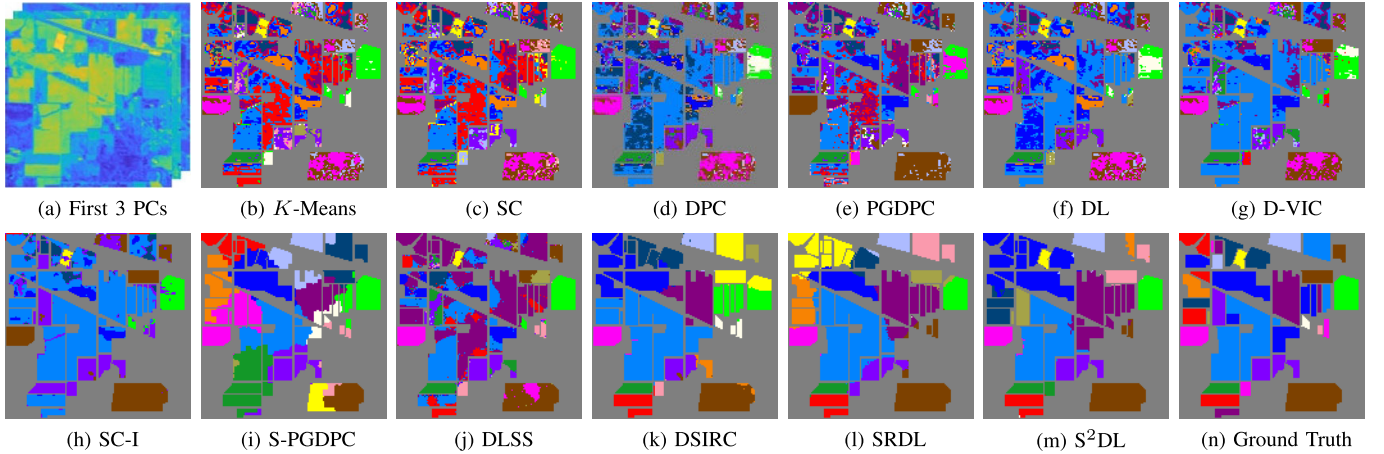


Fig. 3. Comparison of clustering results of (b)–(g) algorithms using spectral information and (h)–(l) algorithms using both spatial and spectral information and (m) S<sup>2</sup>DL on (a) Indian Pines dataset with (n) GT.

diffusion time  $t$ . Given the established stability of DL for the kernel scaling factor  $\sigma_0$  and the number of nearest neighbors  $k_n$  [41], [53], [58], [100], these parameters are not the focus of our robustness analysis here. Instead, we focus on the new hyperparameters introduced with S<sup>2</sup>DL as a part of its incorporation of spatial information and diffusion time.

Fig. 4 illustrates the impact of the number of superpixels, denoted as  $N_s$ , and the spatial radius  $R$  on clustering outcomes ( $k$  held constant). An examination of the algorithm's performance on four HSIs suggests that a larger spatial radius  $R$  is preferable when  $N_s$  is relatively small. This strategy guarantees an adequate pixel count for the construction of a spatially regularized graph. The Salinas dataset, characterized by its expansive size and homogeneous areas, requires a larger spatial radius compared to the other two datasets under consideration, and S<sup>2</sup>DL's performance is fairly consistent

when  $R \in [10, 30]$  and  $N_s \in [300, 1500]$ . On the other hand, the Indian Pines dataset, with its constrained spatial dimensions relative to the Salinas dataset, is more amenable to the selection of a smaller radius. As depicted in Fig. 4(c), a similar result can be achieved when  $R$  falls within  $[5, 15]$  and  $N_s \in [300, 1500]$ . Regarding the Salinas A dataset, its size coupled with consistent spatial clusters make it suitable for a reduced number of superpixels. Notably, S<sup>2</sup>DL can still yield high performance with  $N_s < 1000$ . For the WHU-Hi dataset, the need for a larger spatial radius range of  $R \in [25, 60]$  and a superpixel range of  $N_s \in [300, 1500]$  is attributed to its extensive size. In general, there is an inverse relationship between the number of superpixels  $N_s$  and the spatial radius  $R$  across various datasets. The spatial dimensions and complexity of each dataset also influence the optimal selection of  $N_s$  and  $R$ . Despite these variations, the S<sup>2</sup>DL algorithm demonstrates

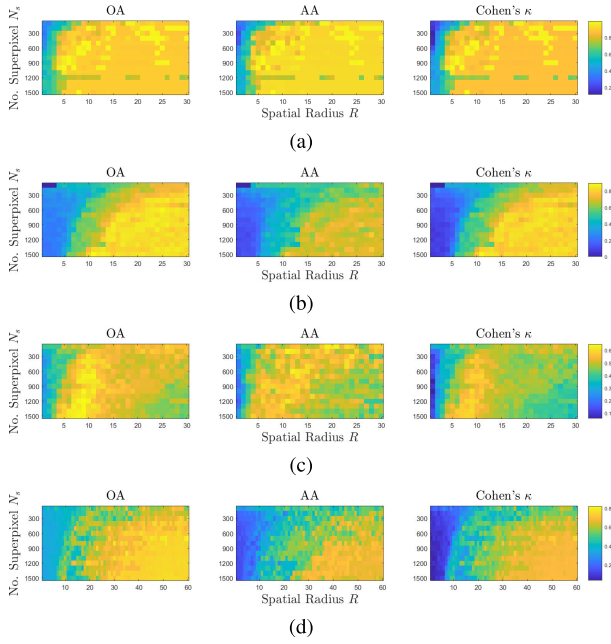


Fig. 4. Analysis of OA, AA, and  $\kappa$  for four HSIs under varying spatial radii  $R$  and number of superpixels  $N_s$ , with a fixed value of  $k$ . Columns in the figure represent the OA, AA, and  $\kappa$ , respectively, while rows correspond to the four different datasets. Each subplot within the figure depicts the performances achieved through various combinations of  $R$  and  $N_s$ . The  $x$ -axis represents the spatial radius, while the  $y$ -axis denotes the number of superpixels. (a) Salinas,  $k = 3$ . (b) Salinas A,  $k = 5$ . (c) Indian Pines,  $k = 5$ . (d) WHU-Hi,  $k = 5$ .

robustness to changes in the hyperparameters  $R$  and  $N_s$ , maintaining consistent performance across HSIs of different sizes and spatial complexities.

Fig. 5 quantifies the robustness of S<sup>2</sup>DL to different spatial radii  $R$  and the number of representative pixels  $k$ . The top row of Fig. 5 shows that, with an increase in spatial radius, the average performance across the four datasets initially experiences a swift uptick before eventually flattening. This pattern not only indicates robustness to changes in the spatial radius but also underscores the enhancement in results upon the integration of spatial information. For the Salinas A and Indian Pines datasets, the flattened phases both occur around  $R = 10$ , while for the Salinas dataset, the flat phase arises near  $R = 20$ . For the WHU-Hi dataset, performance continues to improve with increasing  $R$ , with the gains becoming marginal beyond  $R = 40$ . Similarly, Fig. 5 indicates that S<sup>2</sup>DL is highly robust to the selection of  $k$ —the number of representative pixels sampled from ERS superpixels—across the evaluated four datasets. We observe that S<sup>2</sup>DL achieves the highest performance on Salinas A for  $k > 3$  in Fig. 5. S<sup>2</sup>DL achieves its peak performance on Salinas for  $k = 3$  and has stable performance for  $k > 3$ . Although there is slight fluctuation in the performance of S<sup>2</sup>DL on Indian Pines as  $k$  increases, the variance in peak performance remains consistent within a narrow margin less than 5%. Finally, for the WHU-Hi dataset, performance slightly increases and achieves its best when  $k = 5$ , then slightly drops when  $k = 6$ . This consistent trend across different values of  $k$  demonstrates the robustness of the proposed method for all four datasets under consideration.

Finally, we analyzed the robustness of S<sup>2</sup>DL to the selection of  $t$ , the diffusion time parameter used in diffusion distances, to explicitly understand how the diffusion process influences clustering outcomes. Specifically, we evaluated S<sup>2</sup>DL at the optimal parameter set across a data-dependent exponential grid of  $t$ -values that captures the portion of the diffusion process during which cluster structure may be extracted, see Appendix for more [48], [53], [57]. In Fig. 6, it is evident that, for each dataset, there exists a wide window of diffusion time during which S<sup>2</sup>DL achieves optimal performance. Notably, different algorithms required different diffusion time inputs for optimal clustering performance, likely due to differences in intrinsic geometric structure within the HSIs. That S<sup>2</sup>DL is capable of recovering latent cluster structure during regions of time aligns with the literature in diffusion clustering [42], [48], [57], which has demonstrated that diffusion time is closely linked with the scale of discoverable cluster structure. Notably, this also indicates that S<sup>2</sup>DL may be applied to the important problem of multiscale clustering by varying the diffusion time parameter  $t$  [48], [57], [100]. Regardless, the steady performance overextended periods of diffusion time indicates the algorithm's robustness to this parameter at the scale of interest.

4) *Ablation Studies*: This section provides ablation studies focused on the LBB [95] and the spatial regularization [48], [49] within the S<sup>2</sup>DL algorithm. Our objective is to empirically compare the performance of S<sup>2</sup>DL with and without the implementation of these components. The performance differences are quantitatively analyzed based on the three metrics utilized throughout this section.

Our analysis begins with an examination of LBB, revealing varied impacts on performance across datasets, as depicted in Fig. 7(a). For Salinas, LBB generally enhances performance metrics, although a notable exception occurs at  $k = 1$ , where the limited pixel selection due to the dataset's size causes LBB to overemphasize pixel similarity. Conversely, S<sup>2</sup>DL's performance on Indian Pines is uniform across  $k$  for  $k > 1$ , with or without the use of an LBB. Finally, on Salinas A, incorporating LBB leads to a slight reduction in over half of the metrics assessed, yet its overall performance remains robust. This slight decline can be attributed to the smaller size and the subtle differences between classes of Salinas A; when  $k$  is larger, the inclusion of more noisy pixels can diminish the reliability of LBB.

Next, we turn our attention to the impact of spatial regularization, showcased in Fig. 7(b). This ablation study modifies the S<sup>2</sup>DL by removing the spatial window constraint in the kNN graph construction, thereby allowing the algorithm to consider all representative pixels in its search. For Indian Pines and Salinas, adding spatial regularization leads to improved OA and  $\kappa$ , but at the cost of AA, enhancing overall performance. This indicates spatial regularization tends to favor overall accuracy, possibly at the expense of accurately identifying minor classes. In contrast, Salinas A sees improvements in all metrics with spatial regularization, highlighting its efficacy in focusing searches within spatial neighborhoods for spectral neighbors. Yet, WHU-Hi faces a distinct situation; while AA sees a minor increase, both OA and  $\kappa$  experience notable



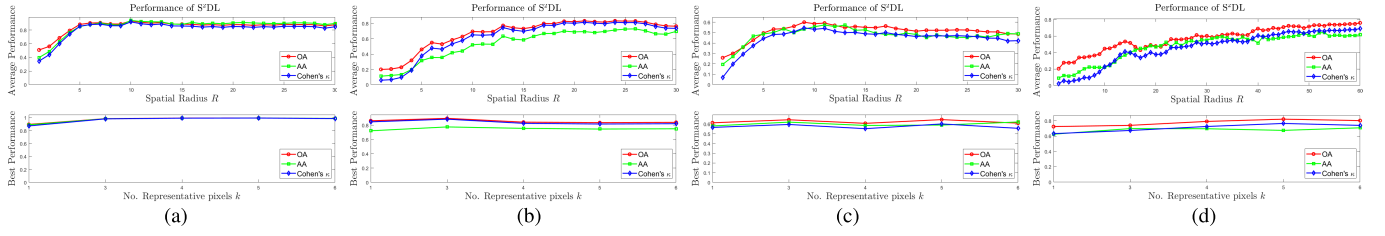


Fig. 5. Average performance in relation to the spatial radius  $R$  (first row) and the best performance concerning the number of representative pixels  $k$  (second row). Each panel corresponds to one of the datasets: (a) Salinas A, (b) Salinas, (c) Indian Pine, and (d) WHU-Hi, with a fixed number of superpixels.

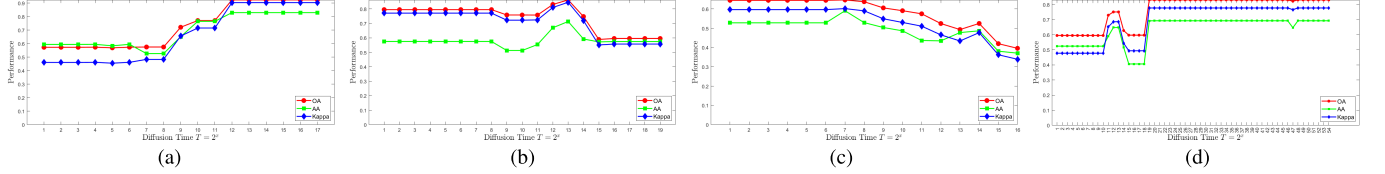


Fig. 6. Analysis of OA, AA, and  $\kappa$  for  $S^2DL$  Across Various Diffusion Times: The x-axis represents the exponential scale of diffusion time (expressed as  $2^x$ ), while the y-axis shows the corresponding values of OA, AA, and  $\kappa$  as the time parameter  $t$  varies across the diffusion process. This figure effectively illustrates the variation and stability of these performance metrics in response to changes in diffusion time. (a) Salinas A. (b) Salinas. (c) Indian Pine. (d) WHU-Hi.

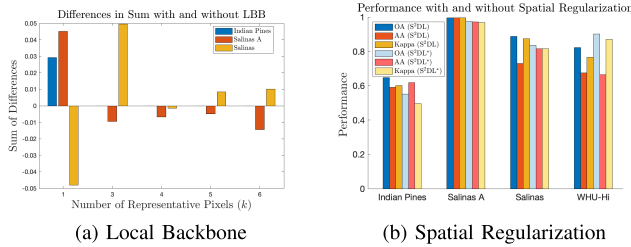


Fig. 7. Performance Impact of LBB and Spatial Regularization in  $S^2DL$ . (a) Summed performance differences across three metrics, calculated as the performance with LBB minus the performance without LBB. (b) Compares the performance metrics between  $S^2DL$  and its variant without spatial regularization,  $S^2DL^*$ .

declines with spatial regularization. The drop is linked to the spatial disconnection of major classes, making local searches ineffective for bridging these areas.

To further investigate the performance on WHU-Hi, we visualize the clusterings produced by algorithms that rely solely on spectral information and demonstrate good performance for WHU-Hi, such as DL and DPC, alongside the clusterings from algorithms that utilize spatially regularized graphs, namely SRDL and  $S^2DL$ , and  $S^2DL$  without spatial regularization ( $S^2DL^*$ ), in Fig. 8. The clusterings by DL and DPC, shown in Fig. 8(a) and (b), although noisy, effectively identify broad-leaf soybean (yellow) and corn (orange) classes. However, SRDL and  $S^2DL$ , shown in Fig. 8(c) and (d), which employ spatial regularization, underperform on these classes due to their spatial disconnection; spatial regularization restricts pixel connections to spectral nearest neighbors within a local spatial domain. Without a sufficiently large radius, establishing connections between pixels of the same class but in disconnected regions becomes challenging.  $S^2DL^*$ , as depicted in Fig. 8(e), significantly improves performance by sacrificing 1% in AA for a 7.9% increase in OA and a 10.4% increase in kappa compared to  $S^2DL$ . In this scenario,  $S^2DL^*$  surpasses all competing algorithms in all metrics, achieving improvements of 4.4% in OA, 12.5% in AA, 6% in  $\kappa$ , and

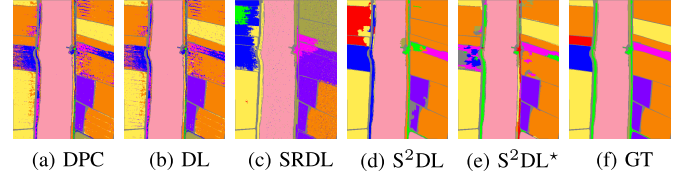


Fig. 8. Comparison of clustering results of (a) and (b) algorithms using spectral information and (c) and (d) algorithms using both spatial and spectral information through spatial regularization and (e) without spatial regularization on (f) WHU-Hi dataset with GT.

a cumulative 22.9% increase in sum, as compared to all algorithms with the exception of  $S^2DL$  itself, as detailed in Table I.

In summary, LBB enhances clustering in spatially homogeneous datasets like Salinas but may be less beneficial for datasets with intricate spatial structures or smaller sizes. Spatial regularization generally enhances clustering performance, yet its application may be counterproductive in cases like WHU-Hi, characterized by spatial disconnection within the same class. The improved performance of  $S^2DL$  without spatial regularization ( $S^2DL^*$ ) on WHU-Hi illustrates the importance of balancing between enhancing connectivity and preserving spatial characteristics.

### B. Experiments on Mangrove Forests in Hong Kong

The MPNR, positioned at the entrance of the Shenzhen River in northwest Hong Kong and facing the Futian Nature Reserve in Shenzhen, spans the coordinates 113°59'E–114°03'E, 22°28'N–22°32'N. This reserve boasts a rich tapestry of ecosystems, including wetlands, freshwater ponds, inter-tidal mudflats, mangroves, reed beds, and fishponds, each fostering a diverse array of wildlife. Recognized for its ecological significance and unique location, MPNR was designated a restricted area in 1975 and subsequently declared a site of special scientific importance in 1976 [51]. At the heart of MPNR lies its expansive mangrove forests, covering

TABLE II

COMPARATIVE PERFORMANCE ANALYSIS OF S<sup>2</sup>DL AND OTHER METHODS ON THE MPNR DATASET. THE TABLE PRESENTS THE OVERALL PERFORMANCE AND PRODUCER'S ACCURACY FOR VARYING  $k$  VALUES IN S<sup>2</sup>DL AND COMPARES IT WITH OTHER CLUSTERING METHODS. BEST PERFORMANCES IN EACH COLUMN FOR BOTH S<sup>2</sup>DL AND OTHER METHODS ARE HIGHLIGHTED IN BOLD

Overall Performance				Producer's Accuracy							
S <sup>2</sup> DL	OA	AA	$\kappa$	Mudflat	Water	KO2	KO1	AM	AI2	AI1	AC
$k = 1$	0.715	0.669	0.664	0.887	<b>0.794</b>	<b>1.000</b>	0.538	0.667	0.917	0.547	0.000
$k = 3$	0.687	0.733	0.637	0.919	0.614	0.808	0.364	0.778	0.875	<b>0.604</b>	<b>0.903</b>
$k = 4$	0.722	0.743	0.673	0.919	0.755	0.808	0.529	0.729	0.875	0.425	<b>0.903</b>
$k = 5$	<b>0.732</b>	<b>0.770</b>	<b>0.686</b>	<b>0.932</b>	0.690	0.808	0.364	<b>0.986</b>	0.875	<b>0.604</b>	<b>0.903</b>
$k = 6$	0.706	0.688	0.652	0.919	0.522	0.808	<b>0.796</b>	0.521	<b>0.938</b>	0.425	0.581
<i>K</i> -Means	0.426	0.331	0.311	0.837	0.190	0	0.378	<b>1.000</b>	0	0	<b>0.226</b>
SC	0.533	0.502	0.452	0.805	0.147	0.635	0.480	0.882	0.844	0.104	0
DPC	0.542	0.463	0.454	0.629	0.489	0	0.578	0.743	0.896	0.142	<b>0.226</b>
PGDPC	0.488	0.357	0.361	0.833	0.288	0	<b>0.991</b>	0	0.521	0	<b>0.226</b>
DL	0.542	0.463	0.454	0.629	0.489	0	0.578	0.743	0.896	0.142	<b>0.226</b>
D-VIC	0.566	0.466	0.476	0.701	0.446	0	0.693	0.826	0.823	0	<b>0.226</b>
SC-I	0.481	0.387	0.375	0.792	0.087	0	0.582	0.896	0.542	0	0.194
S-PGDPC	0.651	0.542	0.586	0.891	0.571	0.019	0.529	0.785	<b>0.958</b>	0.585	0
DLSS	0.592	0.490	0.512	0.805	0.467	0	0.653	0.806	0.927	0.038	<b>0.226</b>
DSIRC	0.549	0.466	0.456	0.842	0.005	0.365	0.680	0.806	0.938	0	<b>0.226</b>
SRDL	<b>0.673</b>	<b>0.626</b>	<b>0.613</b>	<b>0.968</b>	<b>0.636</b>	<b>0.731</b>	0.511	0.431	<b>0.958</b>	<b>0.679</b>	0.097

approximately 319 hectares and recognized as Hong Kong's largest mangrove habitat [101].

Originally, there were six native mangrove species within the Reserve, including *Kandelia obovate* (KO), *Avicennia marina* (AM), *Aegiceras corniculatum* (AC), *Acanthus ilicifolius* (AI), *Bruguiera gymnorhiza* (BG), and *Excoecaria agallocha* (EA). The latter two species are rarer compared to their counterparts. Additionally, two exotic species, *Sonneratia caseolaris* and *Sonneratia apetala*, originating from the nearby Futian Nature Reserve, have been identified within MPNR. These are actively removed to mitigate their potential impact on the native mangrove population. Consequently, four dominant, well-studied species remain: KO, AM, AC, and AI. Specifically, KO and AI exhibit significant intraspecific variation, leading to their classification into sub-species: KO1, KO2, AI1, and AI2 [102].

The hyperspectral data analyzed in this study were acquired using the Advanced Hyperspectral Imaging system aboard the Gaofen-5 Chinese satellite. This dataset encompasses 330 spectral bands, covering a  $92 \times 72$  pixel spatial region with a spatial resolution of 30 m. These bands include both the visible/near-infrared spectrum, with a spectral resolution of 5 nm, and the shortwave infrared spectrum, with a resolution of 10 nm [51]. The study area includes not only the six primary mangrove classes but also classes such as mudflats and water bodies, collectively accounting for a total of 6624 pixels.

The evaluation metrics employed for benchmark HSIs, including the introduction of producer's accuracy, are used in this study. This metric calculates the ratio of correctly classified pixels for a specific class to the total number of GT pixels for that class, thereby assessing class-specific performance. As depicted in Table II, our method demonstrates stable results across different values of the parameter  $k$ , with optimal performance observed at  $k = 5$ , achieving an OA of 0.732, AA of 0.77, and  $\kappa$  of 0.686. Notably, the best performance for half of the six mangrove species (spanning

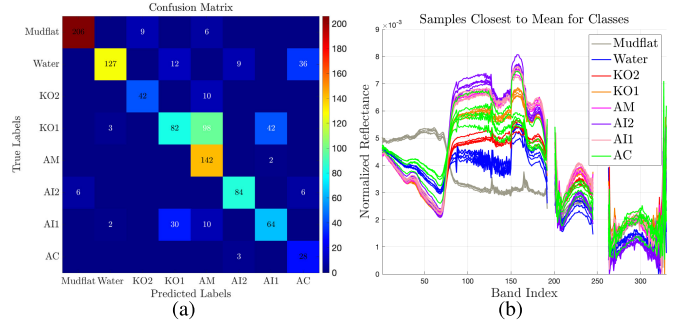


Fig. 9. Confusion matrix for S<sup>2</sup>DL clustering ( $k = 5$ ) and spectral signatures of samples from each class, colored by GT. Notably, different classes exhibit separation within spectral signatures in some spectral bands (but not all), making mangrove species mapping a challenging unsupervised remote sensing problem. (a) Confusion matrix. (b) Spectral signatures.

from column KO2 to AC) is achieved at  $k = 5$ . S<sup>2</sup>DL consistently outperforms other methods in OA, AA, and  $\kappa$ , with SRDL as its nearest competitor. S<sup>2</sup>DL surpasses SRDL in half of the producer accuracy metrics across eight classes, while SRDL faces challenges in precisely identifying the AC and AM classes. In contrast, other methods struggle with accurately clustering classes like KO2, AI1, and AC due to sample imbalance, spectral similarities between classes, and diverse signatures within classes. S<sup>2</sup>DL's success, therefore, highlights the benefit of using representative pixels from each superpixel to reduce the variability within spatial regions prior to cluster analysis.

Fig. 10 presents the visualizations of S<sup>2</sup>DL clustering outcomes for different values of  $k$ . The results are consistently high-performing across all settings, with most classes demonstrating optimal performance at  $k = 5$ . Additionally, Fig. 9(a) presents the confusion matrix for the S<sup>2</sup>DL clustering with  $k = 5$ , where most classes are effectively separated, with notable overlaps in predictions observed for classes such as KO1, AM, and AI1. As depicted in Fig. 9(b), which

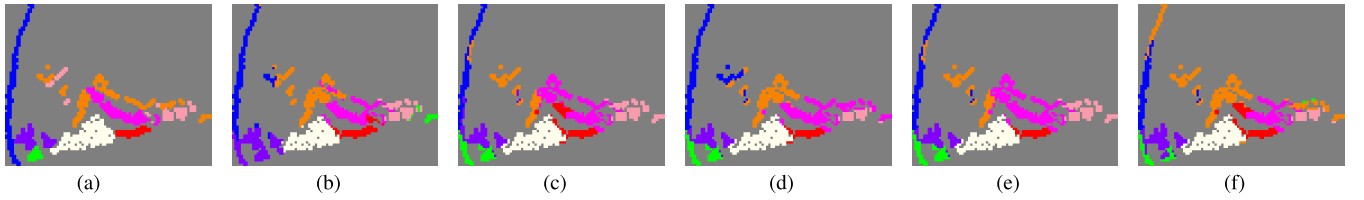


Fig. 10. Comparison of clustering results of  $S^2DL$  on the MPNR dataset with varying number of representative pixels per superpixel  $k$ . The color coding for the GT in (a) is as follows: gray for background, white for mudflat, blue for water, red for KO2, orange for KO1, pink for AM, purple for AI2, light rose for AI1, and green for AC. (a) GT. (b)  $k = 1$ . (c)  $k = 3$ . (d)  $k = 4$ . (e)  $k = 5$ . (f)  $k = 6$ .

showcases the top five samples closest to the mean, AM and AI1 exhibit only subtle spectral differences. In contrast, KO1's spectral signature, which substantially differs from AM and AI1 spectra, leads to the frequent misclassification by  $S^2DL$  of its upper-right and bottom pixels as AM and AI1, respectively, due to the spatial constraints in  $S^2DL$ . The high-quality unsupervised species mappings produced in this section underscore  $S^2DL$ 's capability to deliver robust results in real-world forest environments.

## V. CONCLUSION

This work introduces  $S^2DL$  for unsupervised HSI clustering. Given the high levels of noise and spectral variability often observed in the HSIs, algorithms that rely exclusively on spectral information fail to recover latent cluster structure or produce suboptimal material classifications [2], [4], [45]. To mitigate these important challenges,  $S^2DL$  incorporates both spatial and spectral information, effectively processing HSIs that have consistent spatial patterns but also contain areas with noise or varied spectral characteristics.  $S^2DL$  demonstrates impressive clustering performance on both real-world benchmarking HSIs and the practical landscapes of the MPNR. This balanced performance highlights its robustness in standard scenarios and its adaptability to real-world environmental variations. Moreover,  $S^2DL$  is capable of achieving these high-quality clustering results at a fraction of the computational cost of related algorithms due to its reliance on superpixel segmentation prior to graph construction. Indeed,  $S^2DL$ 's derived superpixels not only encapsulate localized spatial coherence within the image but also reduce the number of data to be analyzed, resulting in a computationally efficient clustering process and a robust utilization of spatial information [39], [89]. Using a spatially regularized graph on this reduced dataset in a diffusion geometry-based clustering procedure enables  $S^2DL$  to efficiently leverage spatial information into a low-RT and highly accurate clustering estimate [48], [49]. Nevertheless, the limitations of the algorithm include dependency on hyperparameter tuning and the method of selecting representative pixels within each superpixel, which might lead to potential information loss.

In future work, we aim to estimate the optimal number of superpixels by leveraging the intrinsic characteristics of datasets, such as size, spatial complexity, and resolution [88], [103]. Since most common superpixel segmentation methods are designed primarily for RGB or grayscale images, they often fall short of fully extracting the abundant spatial and spectral information available in HSIs. Consequently,

exploring and developing superpixel segmentation methods specifically tailored for HSIs will be worthwhile [89], [104]. Additionally, integrating feature extraction techniques into the  $S^2DL$  framework is a promising avenue, allowing us to utilize more effectively the rich spatial and spectral information within superpixels, thereby enhancing the overall performance and efficiency of the algorithm in clustering HSIs [46], [62], [87], [93]. Furthermore, as referenced in Section IV-A3,  $S^2DL$  is expected to be well-equipped to handle multiscale clustering problems by varying its diffusion time parameter. Moreover, by identifying the optimal clustering across scales through minimization of average variation of information [48], [57], [105], we expect to be able to mitigate the dependence of  $S^2DL$  on diffusion time. While this may slightly reduce peak performance, it greatly enhances the practical applicability of the method. Lastly, pursuing the active extension of  $S^2DL$ , especially when a limited number of carefully selected labels are available depending on budget constraints, is a valuable direction for semi-supervised practical applications [41], [50], [71], [106].

## APPENDIX

### OPTIMIZATION OF HYPERPARAMETERS

This appendix details the process by which hyperparameters were tuned in order to obtain the experimental results presented in Section IV. Table III provides a summary of the parameter grids for each algorithm.  $K$ -Means was implemented without the need for hyperparameter adjustments. For stochastic algorithms that require hyperparameter inputs (SC, D-VIC, and DSIRC), optimization was based on achieving the median sum of OA, AA, and  $\kappa$  over ten trials for each set of parameters in the specified hyperparameter grids.

All graph-based algorithms in this study utilized adjacency matrices from sparse kNN graphs, with SC-I employing a full graph. For algorithms without spatial regularization, we used  $\mathcal{N}_1$ , an exponential sampling range of 10–900 for nearest neighbors. For those with spatial regularization,  $\mathcal{N}_2$  was used, with values from 10 to 50 in increments of 10. DL, D-VIC, DLSS, DSIRC, SRDL, and  $S^2DL$  were executed at each  $t$  within  $\mathcal{T} = 0, 1, 2, 2^2, \dots, 2^T$ , where  $T = \lceil \log_2[\log_{\lambda_2(\mathbf{P})}((2 \times 10^{-5})/\min(\pi))] \rceil$ . The process concludes at  $t = 2^T$  since, for  $t \geq 2^T$ ,  $\max_{x,y \in \mathcal{X}} D_t(x, y) \leq 10^{-5}$  [57]. The optimal time step  $t$  from this range was selected for each dataset, maximizing the sum of OA, AA, and  $\kappa$ . Additionally, the KDE and  $\sigma_0$  hyperparameters were uniformly applied across these algorithms. In our grid searches,  $\sigma_0$  covered  $\mathcal{D}$ , which involved sampling  $\ell^2$ -distances between HSI pixels and



TABLE III

HYPERPARAMETER RANGES FOR ALGORITHMS, INCLUDING  $\mathcal{N}_1$  FOR EXPONENTIAL NEAREST NEIGHBOR SAMPLING [10,900],  $\mathcal{N}_2$  FOR SPATIALLY REGULARIZED GRAPHS [10,50],  $\mathcal{D}$  FOR  $\ell^2$ -DISTANCES WITH 1000 NEAREST NEIGHBORS,  $\mathcal{D}_1$  FOR SC-I DISTANCES,  $\mathcal{D}_2$  FOR S-PGDPC GAUSSIAN FILTER  $\sigma$ ,  $\mathcal{T}$  FOR DIFFUSION TIME SAMPLING,  $\mathcal{B}$  FOR SC-I SPATIAL-SPECTRAL INFORMATION RATIO,  $\mathcal{R}$  FOR SPATIAL REGULARIZATION RADII,  $\mathcal{R}_1$  FOR DSIRC ADAPTIVE RADIUS, AND  $\mathcal{S}$  FOR SUPERPIXEL NUMBER RANGE [100–1500]. “—” INDICATES NO HYPERPARAMETER REQUIREMENT

	Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5
K-Means	—	—	—	—	—
SC	$k_n \in \mathcal{N}_1$	—	—	—	—
DPC	$k_n \in \mathcal{N}_1$	$\sigma_0 \in \mathcal{D}$	—	—	—
PGDPC	$k_n \in \mathcal{N}_1$	—	—	—	—
DL	$k_n \in \mathcal{N}_1$	$\sigma_0 \in \mathcal{D}$	$t \in \mathcal{T}$	—	—
D-VIC	$k_n \in \mathcal{N}_1$	$\sigma_0 \in \mathcal{D}$	$t \in \mathcal{T}$	—	—
SC-I	—	$\sigma_1 \in \mathcal{D}_1$	—	$\beta \in \mathcal{B}$	—
S-PGDPC	$k_n \in \mathcal{N}_1$	$\sigma_2 \in \mathcal{D}_2$	—	—	$N_s \in \mathcal{S}$
DLSS	$k_n \in \mathcal{N}_1$	$\sigma_0 \in \mathcal{D}$	$t \in \mathcal{T}$	$R \in \mathcal{R}$	—
DSIRC	$k_n \in \mathcal{N}_1$	$\sigma_0 \in \mathcal{D}$	$t \in \mathcal{T}$	$R_1 \in \mathcal{R}_1$	—
SRDL	$k_n \in \mathcal{N}_2$	$\sigma_0 \in \mathcal{D}$	$t \in \mathcal{T}$	$R \in \mathcal{R}$	—
S <sup>2</sup> DL	$k_n \in \mathcal{N}_2$	$\sigma_0 \in \mathcal{D}$	$t \in \mathcal{T}$	$R \in \mathcal{R}$	$N_s \in \mathcal{S}$

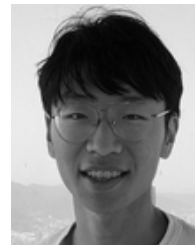
their  $k_n$  nearest neighbors. Additionally,  $\sigma_1$  spanned  $\mathcal{D}_1$  for SC-I, sampling  $\ell^2$ -distances between each data point and all others. For S-PGDPC,  $\sigma_2$  was used as a parameter for Gaussian filtering, applied to blur the image prior to superpixel segmentation. For DLSS, SRDL, and S<sup>2</sup>DL, the spatial radius  $\mathcal{R}$  ranged from 1 to 30, with an exception for the WHU-Hi dataset where this parameter’s range is 1–60, likely due to that dataset’s large spatial dimensions. DSIRC utilized  $\mathcal{R}_1$  to automatically determine the radius of a spatially adaptive window in various directions. SC-I employed  $\mathcal{B}$  as the ratio parameter for balancing spatial and spectral information. The number of superpixels  $N_s$  in  $\mathcal{S}$  was set within a range of 100–1500 in increments of 100.

## REFERENCES

- [1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, “Hyperspectral remote sensing data analysis and future challenges,” *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [2] A. Plaza et al., “Recent advances in techniques for hyperspectral image processing,” *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.
- [3] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, “Deep learning-based classification of hyperspectral data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
- [4] H. Zhai, H. Zhang, P. Li, and L. Zhang, “Hyperspectral image clustering: Current achievements and future lines,” *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 35–67, Dec. 2021.
- [5] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, “Deep learning for hyperspectral image classification: An overview,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
- [6] S. L. Polk, A. H. Y. Chan, K. Cui, R. J. Plemmons, D. A. Coomes, and J. M. Murphy, “Unsupervised detection of ASH dieback disease (*Hymenoscyphus fraxineus*) using diffusion-based hyperspectral image clustering,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2022, pp. 2287–2290.
- [7] R. Li, K. Cui, R. H. Chan, and R. J. Plemmons, “Classification of hyperspectral images using SVM with shape-adaptive reconstruction and smoothed total variation,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2022, pp. 1368–1371.
- [8] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, “A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges,” *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.
- [9] R. A. Borsoi et al., “Spectral variability in hyperspectral data unmixing: A comprehensive review,” *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 223–270, Dec. 2021.
- [10] J. M. Bioucas-Dias et al., “Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [11] R. Heylen, M. Parente, and P. Gader, “A review of nonlinear hyperspectral unmixing methods,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.
- [12] K. Cui and R. J. Plemmons, “Unsupervised classification of AVIRIS-NG hyperspectral images,” in *Proc. 11th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Mar. 2021, pp. 1–5.
- [13] N. Yokoya, C. Grohnfeldt, and J. Chanussot, “Hyperspectral and multispectral data fusion: A comparative review of the recent literature,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.
- [14] L. Loncan et al., “Hyperspectral pansharpening: A review,” *IEEE Geosci. Remote Sens. Mag. Replaces Newsletter*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [15] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, “Deep learning for pixel-level image fusion: Recent advances and future prospects,” *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.
- [16] S. Camalan et al., “Detecting change due to alluvial gold mining in Peruvian rainforest using recursive convolutional neural networks and contrastive learning,” in *Proc. Fall Meeting*, vol. 2022, 2022, p. B52G-0905.
- [17] M. Zhang, W. Li, Y. Zhang, R. Tao, and Q. Du, “Hyperspectral and LiDAR data classification based on structural optimization transmission,” *IEEE Trans. Cybern.*, vol. 53, no. 5, pp. 3153–3164, May 2023.
- [18] Y. Su, L. Gao, M. Jiang, A. Plaza, X. Sun, and B. Zhang, “NSCKL: Normalized spectral clustering with kernel-based learning for semisupervised hyperspectral image classification,” *IEEE Trans. Cybern.*, vol. 53, no. 10, pp. 6649–6662, Oct. 2023.
- [19] R. H. Chan and R. Li, “A 3-stage spectral-spatial method for hyperspectral image classification,” *Remote Sens.*, vol. 14, no. 16, p. 3998, Aug. 2022.
- [20] K. Cui et al., “Semi-supervised change detection of small water bodies using RGB and multispectral images in Peruvian rainforests,” in *Proc. 12th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Sep. 2022, pp. 1–5.
- [21] H. Liu, W. Li, X.-G. Xia, M. Zhang, C.-Z. Gao, and R. Tao, “Central attention network for hyperspectral imagery classification,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8989–9003, Nov. 2023.
- [22] H. Liu, W. Li, X.-G. Xia, M. Zhang, and R. Tao, “Multiarea target attention for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5524916.
- [23] Y. Duan, F. Luo, M. Fu, Y. Niu, and X. Gong, “Classification via structure-preserved hypergraph convolution network for hyperspectral image,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5507113.
- [24] S. Camalan et al., “Change detection of Amazonian alluvial gold mining using deep learning and Sentinel-2 imagery,” *Remote Sens.*, vol. 14, no. 7, p. 1746, Apr. 2022.
- [25] T. Guo, R. Wang, F. Luo, X. Gong, L. Zhang, and X. Gao, “Dual-view spectral and global spatial feature fusion network for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5512913.
- [26] H. Zhou, F. Luo, H. Zhuang, Z. Weng, X. Gong, and Z. Lin, “Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5508614.
- [27] M. Zhang, W. Li, X. Zhao, H. Liu, R. Tao, and Q. Du, “Morphological transformation and spatial-logical aggregation for tree species classification using hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501212.
- [28] M. Zhang, X. Zhao, W. Li, Y. Zhang, R. Tao, and Q. Du, “Cross-scene joint classification of multisource data with multi-level domain adaption network,” *IEEE Trans. Neural Netw. Learn. Syst.*, early access, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10093906>
- [29] H. Zhao, F. Zhou, L. Bruzzone, R. Guan, and C. Yang, “Superpixel-level global and local similarity graph-based clustering for large hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5519316.

- [30] Y. Ding et al., "Self-supervised locality preserving low-pass graph convolutional embedding for large-scale hyperspectral image clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536016.
- [31] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.
- [32] Y. Cai, Z. Zhang, Z. Cai, X. Liu, X. Jiang, and Q. Yan, "Graph convolutional subspace clustering: A robust subspace clustering framework for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4191–4202, May 2021.
- [33] B. Peng, Y. Yao, J. Lei, L. Fang, and Q. Huang, "Graph-based structural deep spectral-spatial clustering for hyperspectral image," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 5502112.
- [34] S. Huang, H. Zhang, and A. Pizurica, "Hybrid-hypergraph regularized multiview subspace clustering for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5505816.
- [35] H. Zhai, H. Zhang, L. Zhang, P. Li, and A. Plaza, "A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 43–47, Jan. 2017.
- [36] R. Wang, F. Nie, and W. Yu, "Fast spectral clustering with anchor graph for large hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2003–2007, Nov. 2017.
- [37] Y. Zhao, Y. Yuan, and Q. Wang, "Fast spectral clustering for unsupervised hyperspectral image classification," *Remote Sens.*, vol. 11, no. 4, p. 399, Feb. 2019.
- [38] Q. Wang, Y. Miao, M. Chen, and Y. Yuan, "Spatial-spectral clustering with anchor graph for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5542413.
- [39] X. Chen, Y. Zhang, X. Feng, X. Jiang, and Z. Cai, "Spectral-spatial superpixel anchor graph-based clustering for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5507405.
- [40] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, Mar. 2005.
- [41] J. M. Murphy and M. Maggioni, "Unsupervised clustering and active learning of hyperspectral images with nonlinear diffusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1829–1845, Mar. 2019.
- [42] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, Jul. 2006.
- [43] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, "Noise reduction in hyperspectral images: Overview and application," *Remote Sens.*, vol. 10, no. 3, p. 482, Mar. 2018.
- [44] W. Cai, J. Jiang, and J. Qian, "Large-scale hyperspectral image restoration via a superpixel distributed algorithm based on graph signal processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5503517.
- [45] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
- [46] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.
- [47] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.
- [48] S. L. Polk and J. M. Murphy, "Multiscale clustering of hyperspectral images through spectral-spatial diffusion geometry," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 4688–4691.
- [49] J. M. Murphy and M. Maggioni, "Spectral-spatial diffusion geometry for hyperspectral image clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1243–1247, Jul. 2020.
- [50] J. M. Murphy, "Spatially regularized active diffusion learning for high-dimensional images," *Pattern Recognit. Lett.*, vol. 135, pp. 213–220, Jul. 2020.
- [51] L. Wan, Y. Lin, H. Zhang, F. Wang, M. Liu, and H. Lin, "GF-5 hyperspectral data for species mapping of mangrove in Mai Po, Hong Kong," *Remote Sens.*, vol. 12, no. 4, p. 656, Feb. 2020.
- [52] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics), vol. 2. New York, NY, USA: Springer, 2009.
- [53] S. L. Polk, K. Cui, A. H. Y. Chan, D. A. Coomes, R. J. Plemmons, and J. M. Murphy, "Unsupervised diffusion and volume maximization-based clustering of hyperspectral images," *Remote Sens.*, vol. 15, no. 4, p. 1053, Feb. 2023.
- [54] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, no. 14, Oakland, CA, USA, 1967, pp. 281–297.
- [55] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annu. Rev. Statist. Appl.*, vol. 6, pp. 355–378, Jan. 2019.
- [56] M. Ester et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, 1996, vol. 96, no. 34, pp. 226–231.
- [57] J. M. Murphy and S. L. Polk, "A multiscale environment for learning by diffusion," *Appl. Comput. Harmon. Anal.*, vol. 57, pp. 58–100, Mar. 2022.
- [58] M. Maggioni and J. M. Murphy, "Learning by unsupervised nonlinear diffusion," *J. Mach. Learn. Res.*, vol. 20, no. 160, pp. 1–56, 2019.
- [59] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
- [60] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long, "A survey of clustering with deep learning: From the perspective of network architecture," *IEEE Access*, vol. 6, pp. 39501–39514, 2018.
- [61] Y. Cai et al., "Large-scale hyperspectral image clustering using contrastive learning," 2021, *arXiv:2111.07945*.
- [62] Y. Cai et al., "Superpixel contracted neighborhood contrastive subspace clustering network for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5530113.
- [63] Y. Ding et al., "Unsupervised self-correlated learning smoothy enhanced locality preserving graph convolution embedding clustering for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536716.
- [64] K. Li, Y. Qin, Q. Ling, Y. Wang, Z. Lin, and W. An, "Self-supervised deep subspace clustering for hyperspectral images with adaptive self-expressive coefficient matrix initialization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3215–3227, 2021.
- [65] Y. Zhao and X. Li, "Deep spectral clustering with regularized linear embedding for hyperspectral image clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509311.
- [66] D. Bandyopadhyay, S. Mukherjee, J. Ball, G. Vincent, D. A. Coomes, and C.-B. Schönlieb, "Tree species classification from hyperspectral data using graph-regularized neural networks," 2022, *arXiv:2208.08675*.
- [67] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *Proc. CVPR*, Jun. 2015, pp. 427–436.
- [68] C. Szegedy et al., "Intriguing properties of neural networks," 2013, *arXiv:1312.6199*.
- [69] B. D. Haeffele, C. You, and R. Vidal, "A critique of self-expressive deep subspace clustering," 2020, *arXiv:2010.03697*.
- [70] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. NeurIPS*, vol. 14, 2001, pp. 1–8.
- [71] S. L. Polk, K. Cui, R. J. Plemmons, and J. M. Murphy, "Active diffusion and VCA-assisted image segmentation of hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2022, pp. 1364–1367.
- [72] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, "Extended random Walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.
- [73] X. Zhao et al., "Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020.
- [74] B. Cui, X. Xie, X. Ma, G. Ren, and Y. Ma, "Superpixel-based extended random Walker for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3233–3243, Jun. 2018.
- [75] K. Cui, R. Li, S. L. Polk, J. M. Murphy, R. J. Plemmons, and R. H. Chan, "Unsupervised spatial-spectral hyperspectral image reconstruction and clustering with diffusion geometry," in *Proc. 12th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Sep. 2022, pp. 1–5.
- [76] X. Yang, Y. Xu, S. Li, Y. Liu, and Y. Liu, "Fuzzy embedded clustering based on bipartite graph for large-scale hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5505605.
- [77] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. CVPR*, Jun. 2011, pp. 2097–2104.
- [78] M. Wang, X. Liu, Y. Gao, X. Ma, and N. Q. Soomro, "Superpixel segmentation: A benchmark," *Signal Process., Image Commun.*, vol. 56, pp. 28–39, Aug. 2017.

- [79] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [80] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [81] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Proc. ECCV*, D. Forsyth, P. Torr, and A. Zisserman, Eds. Berlin, Germany: Springer, 2008, pp. 705–718.
- [82] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [83] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
- [84] D. Tang, H. Fu, and X. Cao, "Topology preserved regular superpixel," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 765–768.
- [85] Y.-J. Liu, M. Yu, B.-J. Li, and Y. He, "Intrinsic manifold SLIC: A simple and efficient method for computing content-sensitive superpixels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 653–666, Mar. 2018.
- [86] X. Xie, G. Xie, X. Xu, L. Cui, and J. Ren, "Automatic image segmentation with superpixels and image-level labels," *IEEE Access*, vol. 7, pp. 10999–11009, 2019.
- [87] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.
- [88] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
- [89] P. Sellars, A. I. Aviles-Rivero, and C.-B. Schönlieb, "Superpixel contracted graph-based learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4180–4193, Jun. 2020.
- [90] C. Zhao et al., "Hyperspectral image classification with multi-attention transformer and adaptive superpixel segmentation-based active learning," *IEEE Trans. Image Process.*, vol. 32, pp. 3606–3621, 2023.
- [91] H. Liu, W. Li, X.-G. Xia, M. Zhang, and R. Tao, "Superpixelwise collaborative-representation graph embedding for unsupervised dimension reduction in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4684–4698, 2021.
- [92] J. Mei et al., "PSASL: Pixel-level and superpixel-level aware subspace learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4278–4293, Jul. 2019.
- [93] Z. Chen, J. Jiang, C. Zhou, S. Fu, and Z. Cai, "SuperBF: Superpixel-based bilateral filtering algorithm and its application in feature extraction of hyperspectral images," *IEEE Access*, vol. 7, pp. 147796–147807, 2019.
- [94] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—I," in *Mathematical Programming*, vol. 14. 1978, pp. 265–294.
- [95] S. A. Seyed, A. Lotfi, P. Moradi, and N. N. Qader, "Dynamic graph-based label propagation for density peaks clustering," *Exp. Syst. Appl.*, vol. 115, pp. 314–328, Jan. 2019.
- [96] A. Beygelzimer, S. Kakade, and J. Langford, "Cover trees for nearest neighbor," in *Proc. ICML*, 2006, pp. 97–104.
- [97] J. Guan, S. Li, X. He, and J. Chen, "Peak-graph-based fast density peak clustering for image segmentation," *IEEE Signal Process. Lett.*, vol. 28, pp. 897–901, 2021.
- [98] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.
- [99] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, Mar. 1957.
- [100] S. L. Polk, "Diffusion-based clustering of high-dimensional datasets," Ph.D. dissertation, Dept. Math., Tufts Univ., Medford, MA, USA, 2022.
- [101] M. Jia, Y. Zhang, Z. Wang, K. Song, and C. Ren, "Mapping the distribution of mangrove species in the Core Zone of Mai Po Marshes Nature Reserve, Hong Kong, using hyperspectral data and high-resolution data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 33, pp. 226–231, Dec. 2014.
- [102] T. Wang, H. Zhang, H. Lin, and C. Fang, "Textural–spectral feature-based species classification of mangroves in Mai Po Nature Reserve from worldview-3 imagery," *Remote Sens.*, vol. 8, no. 1, p. 24, 2016.
- [103] B. Tu, W. Kuang, G. Zhao, and H. Fei, "Hyperspectral image classification via superpixel spectral metrics representation," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1520–1524, Oct. 2018.
- [104] M. P. Barbato, P. Napoletano, F. Piccoli, and R. Schettini, "Unsupervised segmentation of hyperspectral remote sensing images with superpixels," *Remote Sens. Appl., Soc. Environ.*, vol. 28, Nov. 2022, Art. no. 100823.
- [105] M. Meilä, "Comparing clusterings—An information based distance," *J. Multivariate Anal.*, vol. 98, no. 5, pp. 873–895, 2007.
- [106] M. Maggioni and J. M. Murphy, "Learning by active nonlinear diffusion," *Found. Data Sci.*, vol. 1, no. 3, pp. 271–291, 2019.



**Kangning Cui** received the B.S. degree in computing mathematics from the City University of Hong Kong, Hong Kong, in 2021. He is currently pursuing the Ph.D. degree in mathematics with the City University of Hong Kong, and the Hong Kong Center for Cerebro-Cardiovascular Health Engineering, Hong Kong.

He has been researched with the Department of Computer Science, Wake Forest University, Winston-Salem, NC, USA, since 2020. His research interests include image processing and machine learning.



**Ruoning Li** received the B.S. degree in applied mathematics from Harbin Institute of Technology, Weihai, China, in 2020. She is currently pursuing the Ph.D. degree in mathematics with the City University of Hong Kong, Hong Kong.

Her research has included collaboration with Wake Forest University, Winston-Salem, NC, USA. Her research interests include image processing, signal processing, and machine learning.



**Sam L. Polk** received the B.S. degree in mathematics and the B.A. degree in economics from the University of Rochester, Rochester, NY, USA, in 2017, the M.S. degree in mathematics from Tufts University, Medford, MA, USA, in 2020, and the Ph.D. degree in mathematics from Tufts University, in 2022, for research on unsupervised diffusion-based hyperspectral image clustering algorithms.

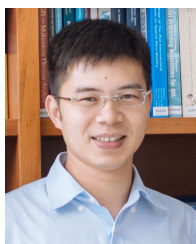
From 2018 to 2019, he researched with the Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA. Since 2021, he has researched machine learning foundations with the MIT Lincoln Laboratory, Lexington, MA, USA. His research interests include diffusion geometry, sequential autoencoding, network science, unsupervised machine learning, and image processing.





**Yinyi Lin** received the B.Eng. degree in geographical conditions monitoring from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2017, and the Ph.D. degree in earth system and geoinformation science from The Chinese University of Hong Kong, Hong Kong, in 2021.

She is currently a Post-Doctoral Fellow with the Department of Geography, The University of Hong Kong, Hong Kong. Her research interests include multisource data fusion and urban remote sensing.



**Hongsheng Zhang** received the B.Eng. degree in computer science and technology and the M.Eng. degree in computer applications technology from South China Normal University, Guangzhou, China, in 2007 and 2010, respectively, and the Ph.D. degree in earth system and geoinformation science from The Chinese University of Hong Kong, Hong Kong, in 2013.

He is currently an Assistant Professor with the Department of Geography, The University of Hong Kong, Hong Kong. His research is mainly focused

on incorporating multiple remote sensing technologies to understand the urbanization process and its environmental and ecological impacts at various spatial and temporal scales.



**James M. Murphy** received the B.S. degree in mathematics from the University of Chicago, Chicago, IL, USA, in 2011, and the Ph.D. degree in mathematics from the University of Maryland, College Park, MD, USA, in 2015.

He was an Assistant Research Professor with the Department of Mathematics, Duke University, Durham, NC, USA, from 2015 to 2016, and an Assistant Research Scientist and a Senior Lecturer with the Department of Mathematics, Johns Hopkins University, Baltimore, MD, USA,

from 2016 to 2018. He is currently an Assistant Professor of mathematics at Tufts University, Medford, MA, USA. His research interests include applied harmonic analysis, machine and statistical learning, graph theory, high-dimensional probability and statistics, and image, signal, and network processing.



**Robert J. Plemmons** received the B.S. degree in mathematics and physics from Wake Forest University, Winston-Salem, NC, USA, in 1961, and the Ph.D. degree in applied mathematics from Auburn University, Auburn, AL, USA, in 1965.

From 1961 to 1965, he was a Professional Baseball Player with Baltimore Orioles Farm System, Baltimore, Maryland. In 1965, he was a Research Engineer with Martin Aerospace Company, Orlando, Florida. From 1965 to 1967, he was a Research Scientist with the National Security Agency, Fort Meade, Maryland. From 1967 to 1981, he was with the Department of Mathematics and the Department of Computer Science, The University of Tennessee, Knoxville, TN, USA, where he was an Associate Professor and then a Professor. From 1981 to 1991, he was a Professor with the Department of Mathematics and the Department of Computer Science, North Carolina State University, Raleigh, NC, USA. He was a Visiting Professor at Stanford University, Stanford, CA, USA, in 1978; the University of Illinois, Urbana-Champaign, Champaign, IL, USA, in 1986; the University of Minnesota, Minneapolis, MN, USA, in 1992; and Duke University, Durham, NC, USA, in 1998. Since 1991, he has been a Z. Smith Reynolds Professor with the Departments of Mathematics and Computer Science, Wake Forest University. His research interests include scientific computation, numerical linear algebra, ill-posed inverse problems, image processing, and hyperspectral and LiDAR methods in remote sensing.



**Raymond H. Chan** received the B.Sc. degree in mathematics from The Chinese University of Hong Kong, Hong Kong, in 1980, and the M.Sc. and Ph.D. degrees in applied mathematics from the Courant Institute of Mathematical Sciences, New York University, New York, NY, USA, in 1984 and 1985, respectively.

He is currently the Dean of the College of Science, City University of Hong Kong, Hong Kong, and the Co-Director of the Hong Kong Center for Cerebro-Cardiovascular Health Engineering, Hong Kong. His research interests include scientific computation, numerical linear algebra, and image processing problems.