Hypergraph Convolutional Subspace Clustering With Multihop Aggregation for Hyperspectral Image

Zijia Zhang, Yaoming Cai, *Student Member, IEEE*, Wenyin Gong, *Member, IEEE*, Pedram Ghamisi, *Senior Member, IEEE*, Xiaobo Liu, *Senior Member, IEEE*, and Richard Gloaguen

***Abstract*—Subspace clustering methods have become a powerful tool to cluster hyperspectral imaging (HSI) data as they ensure theoretical guarantees and empirical success. However, existing methods explore subspace representation in the Euclidean space, and thus, failing to exploit the high-order relationship and long-range interdependences. This article presents a simple yet effective method, to extend subspace clustering into the non-Euclidean domain entitled hypergraph convolutional subspace clustering (HGCSC). Instead of treating HSI as Euclidean data only, we represent all the intraclass relations as hyperedges in a hypergraph. With this representation, we can recast the classic self-expression as a hypergraph convolutional self-representation model. To explore the long-range neighboring relation, we introduce a multihop hypergraph convolution process into the method by collapsing the repeated multiplications into a single matrix. HGCSC adopts the Frobenius norm to ensure a closed-form solution. We assess the performance of HGCSC on five real HSI datasets and show that HGCSC significantly outperforms competitors in terms of clustering accuracy.**

***Index Terms*—Graph representation learning, hypergraph convolutional networks (HGCN), hyperspectral image clustering, subspace clustering.**

I. Introduction

HYPERSPECTRAL imaging (HSI) is one of the most important Earth observation (EO) technologies [1]. The imaging data acquired by hyperspectral sensors consist of dozens or hundreds of narrow spectral bands, which provide rich spectral and spatial information. Methods aiming to automatically find useful information in raw HSI data have increasingly become an active area of research, in EO [2]–[4] and biomedical engineering [5].

Despite the explosive growth of HSI classification methods [6], most of them are supervised, and thus, depend on the availability of labeled samples, which demand not only a large amount of manpower but also limit their applicability [7], [8]. Unsupervised HSI classification (a.k.a, HSI clustering) is a promising paradigm to circumvent human annotation. Unfortunately, due to the complexity and the ubiquitous spectral variability of HSIs [9], HSI clustering is often more challenging than that of RGB images and multispectral images [10]. To this date, there is still a noticeable gap in terms of accuracy between supervised and unsupervised methods [11].

Recently, there has been an increasing interest in narrowing this gap in the hyperspectral community. Classic clustering baselines including K-Means [12], fuzzy c means (FCM) [13], and spectral clustering (SC) [14] have been proven to be adaptable to HSI. However, these methods usually rely on similarity measurement and are not sufficient to explore high-order interactions in the data. Benefiting from the development of representation learning, many advanced clustering methods have recently been devised. One of the most representative models is subspace clustering [15]–[17]. Such methods are based on expressing each data value as a linear combination of other values and have become a popular tool to cluster HSI due to their empirical success and theoretical guarantees.

The key behind subspace clustering is to calculate the self-expression coefficient matrix. Previous works have often assumed that the matrix is sparse (e.g., sparse subspace clustering (SSC) [18]), dense (e.g., efficient dense subspace clustering (EDSC) [19]), low-rank (e.g., low-rank subspace clustering (LRSC) [20]), block diagonal [21], and so on. Furthermore, hybrid variants of these assumptions are often considered in practice [22], [23]. Despite great success, classic subspace clustering models suffer from poor robustness since they are commonly defined in the raw feature space. To overcome this drawback, many works have attempted to explore the self-expressiveness in robust spaces including deep latent space and non-Euclidean space. Pan *et al.* [24] proposed to learn the self-expressiveness in a multilayer autoencoder, resulting in a deep subspace clustering (DSC) model. Such a model has been successfully applied into HSI in [25] and achieved promising results. Furthermore, Cai *et al.* [7] presented a graph convolutional subspace clustering (GCSC) framework to generalize the subspace clustering model into an irregular space. Unlike the graph regularization

Manuscript received September 30, 2021; revised November 18, 2021 and December 9, 2021; accepted December 12, 2021. Date of publication December 21, 2021; date of current version January 7, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62076225 and Grant 62073300, in part by the Natural Science Foundation for Distinguished Young Scholars of Hubei under Grant 2019CFA081, and in part by the Fundamental Research Funds for National Universities, China University of Geosciences (Wuhan). *(Corresponding author: Wenyin Gong.)*

Zijia Zhang and Yaoming Cai are with the School of Computer Science, China University of Geosciences, Wuhan 430074, China, and also with Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology, 09599 Freiberg, Germany (e-mail: zhangzijia@cug.edu.cn; caiyaom@cug.edu.cn).

Wenyin Gong is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: wygong@cug.edu.cn).

Pedram Ghamisi is with the Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology, 09599 Freiberg, Germany, and also with the Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna 1030, Austria (e-mail: p.ghamisi@gmail.com).

Xiaobo Liu is with the School of Automation, China University of Geosciences, Wuhan 430074, China, and also with the Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, China University of Geosciences, Wuhan 430074, China (e-mail: xbliu@cug.edu.cn).

Richard Gloaguen is with the Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology, 09599 Freiberg, Germany (e-mail: r.gloaguen@hzdr.de).

Digital Object Identifier 10.1109/JSTARS.2021.3136599

[26], [27], GCSC provides a more efficient manner to incorporate structure with feature information from the perspective of graph representation learning. However, a simple graph structure is unable to reveal the high-order relationship of data.

The recent development of generalizing convolution on graphs [28], [29] has bridged deep learning and structural data. It also enables us to revisit previous problems from a non-Euclidean space, even though there are no explicit structures. By treating HSI as a single graph or a set of graphs, graph neural networks (GNNs) have been widely used for supervised [30] and semisupervised [31] classification of HSI. The existing models follow a message passing scheme by aggregating neighboring information from either spatial domain [32] or spectral domain [31], making it efficient in capturing more complex relationships beyond the original spectral-spatial information in HSI. However, only a few attempts have been made to incorporate GNNs into unsupervised HSI classification.

In this article, we present an improved GCSC approach and consider the complex data correlation in HSI clustering. We term the proposed approach hypergraph convolutional subspace clustering (HGCSC). Confronting the challenges of learning self-expressive coefficients for complex HSI, we propose to incorporate the data structure in a hypergraph, which is more flexible for subspace modeling, especially when dealing with intraclass variation. Technically, HGCSC follows the basic GCSC framework but generalizes it by using hypergraph convolution. Furthermore, we introduce a simplified multihop propagation mechanism to ensure the incorporation of long-range neighboring relationships in the framework.

To sum up, the contribution of this article is twofold. First, by generalizing subspace clustering into a hypergraph, we propose a simple but effective method with multihop neighbor aggregation for HSI clustering. We train HGCSC by calculating the closed-form solution induced by the Frobenius norm, resulting in a high computational efficiency. Second, we systematically assess the quality and effectiveness of HGCSC on real HSI datasets, showing that HGCSC is consistently superior to baselines by a significant margin across all datasets.

The rest of this article is organized as follows. We first briefly review subspace clustering and introduce hypergraph concepts in Section II. Second, we introduce the details of the proposed HGCSC method in Section III. Finally, in Section IV, we systematically qualitatively and quantitively assess the proposed method. Finally, Section V concludes this article.

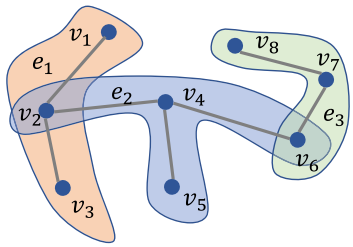
II. Preliminaries

A. Notations

Throughout this article, we use boldface lowercase italics (e.g., $\boldsymbol{x}$), uppercase roman symbols (e.g., $\mathbf{X}$), regular italics (e.g., x_{ij}), and calligraphy symbols (e.g., $\mathcal{T}$) to indicate vectors, matrices, scalars, and sets, respectively. The Frobenius norm of a matrix $\mathbf{X}$ is defined as $\|\mathbf{X}\|_F = \sqrt{tr(\mathbf{X}\mathbf{X}^T)}$. A graph structure is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where $\mathcal{V}$ denotes the node set of the graph with $v_i \in \mathcal{V}$ and $|\mathcal{V}| = N$, $\mathcal{E}$ indicates the edge set with $(v_i, v_j) \in \mathcal{E}$, and $\mathbf{A} \in \mathbb{R}^{N\times N}$ stands for the adjacency matrix of the graph.

B. Hypergraph Learning

Hypergraph is the generalization of the simple graph, in which its edges, i.e., hyperedges can join any number of vertices.



(a)

	e_1	e_2	e_3
v_1	1	0	0
v_2	1	1	0
v_3	1	0	0
v_4	0	1	0
v_5	0	1	0
v_6	0	1	1
v_7	0	0	1
v_8	0	0	1

(b)

Fig. 1. Example of (a) hypergraph consisting of eight vertices and three hyperedges and (b) its corresponding incidence matrix $\mathbf{H}$ in $\mathbb{R}^{8\times3}$.

Fig. 1 shows a simple example of hypergraph. Formally, a hypergraph with N vertices and M hyperedges is represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V}$ denotes the vertex set with $v_i \in \mathcal{V}$, $\mathcal{E}$ indicates the hyperedge set with $e \in \mathcal{E}$, and $w(e) \in \mathcal{W}$ stands for the weight of hyperedge e. The hypergraph $\mathcal{G}$ is often expressed in an incidence matrix $\mathbf{H} \in \mathbb{R}^{N\times M}$, where $h(v, e)$ denotes the connection of v and e and can be determined by

$$h(v,e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{otherwise} \end{cases}. \tag{1}$$

The degree of a vertex v is defined as $d(v) = \sum_{e\in\mathcal{E}} w(e)h(v,e)$, and the degree of an hyperedge e is defined by $d(e) = \sum_{v\in\mathcal{V}} h(v,e)$. For each vertex and hyperedge, their diagonal degree matrices $\mathbf{D}_v \in \mathbb{R}^{N\times N}$ and $\mathbf{D}_e \in \mathbb{R}^{M\times M}$ are, respectively, defined as $[\mathbf{D}_v]_{ii} = \sum_e^M w(e)H_{ie}$ and $[\mathbf{D}_e]_{ii} = \sum_i^N H_{ie}$. Let $\mathbf{W} \in \mathbb{R}^{M\times M}$ indicates the diagonal matrix of edge weights. A hypergraph learning task often involves a normalized hypergraph Laplacian matrix [33], [34] defined by

$$\mathbf{L} = \mathbf{I} - \mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_v^{-1/2}. \tag{2}$$

Practically, hypergraphs are usually used as an alternative to a simple graph, such as hypergraph regularizations in semisupervised [11] and unsupervised learning problems [35].

Compared to the simple graph, which can only represent one-to-one relationships, hypergraph can be used to indicate more complex relationships, e.g., one-to-many and many-to-many relationships. For a classification task, the intraclass variations between data points can be naturally defined as hyperedges of a hypergraph, which often results in a more robust result. More recently, in analogy to graph convolutional networks [28], hypergraph has been used to defined convolution (i.e., hypergraph convolutional networks (HGCN) [36], [37]) for complex structural data.

C. HSI Clustering

HSI clustering aims to group every HSI cells into several distinct clusters. Due to the lack of human annotations, such a task is essentially a typical NP-hard (nondeterministic polynomial) combinatorial optimization problem. Luckily, there are often many intrinsic clues about HSI values belonging to the same cluster, making the problem solvable. Nonetheless, HSI clustering is still more challenging than supervised HSI classification problems because of the difficulty to assess all the parameters that define classes without prior knowledge [10].

The core of HSI clustering is to find out the criteria for the assignment to a specific label. Classic centroid-based clustering models, e.g., K-Means [12] and fuzzy c means (FCM) [13], frequently adopt similarity measurements, such as the Euclidean distance as the criterion. Recently, subspace clustering has achieved great success in various areas. For HSI clustering, many promising variants have been presented, such as spectral-spatial sparse subspace clustering (S^4C) [18], kernel SSC [38], and multiobjective-based SSC [39].

On the other hands, this type of methods often suffers from poor robustness to noise and outliers and high sensitivity to the choice of cluster centroids. To overcome the drawbacks, many attempts have been devoted to designing clustering models by using the spectral graph theory. Such models recast clustering problems into a graph partitioning (e.g., SC [14], [40]), resulting in two benefits: good theoretical guarantees and excellent robustness to noise.

With the development of deep learning, there is increasing interest in deep clustering for HSI. For example, DSC [24], [41] has been proven to be superior to the traditional shallow subspace clustering models. Such a method aims to explore the subspace representation coefficients in a nonlinear latent space approximated by an autoencoder, wherein the subspace clustering model will be recast as a special self-expression layer [41]. By incorporating suitable priors, DSC methods can achieve very reliable clustering performance [11], [25]. Furthermore, recent techniques including graph convolution [7] and self-supervised learning [42] have shown some encouraging directions for HSI clustering.

D. Subspace Clustering Models

Let $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{D\times N}$ be a collection of N data points $\{\boldsymbol{x}_i \in \mathbb{R}^D\}_{i=1}^N$ drawn from a union of n subspaces $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \ldots \cup \mathcal{S}_n$, where N and D denote the number of data points and features, respectively. Subspace clustering model follows the reasonable assumption that each data point can be expressed as a linear combination of a few other data points. The observation can be further formulated as the following self-representation problem [16], i.e.,

$$\min \|\mathbf{Z}\|_p \,\text{s.t.}\, \mathbf{X}\mathbf{Z} = \mathbf{X},\ \text{s.t. diag}\,(\mathbf{Z}) = 0. \tag{3}$$

Here, $\mathbf{Z} \in \mathbb{R}^{N\times N}$ denotes the self-expressive coefficient matrix, and $\|\mathbf{Z}\|_p$ represents a p-norm of the matrix $\mathbf{Z}$. Generally, l_1-norm ($\|\mathbf{Z}\|_1$) tends to imply a sparse coefficient matrix, e.g., SSC [16], while nuclear norm ($\|\mathbf{Z}\|_*$) can lead to a low-rank coefficient matrix, e.g., LRSC [20], [40]. The self-representation problem can be solved by using alternating directions method of multipliers [18], [40] and heuristic methods [39], [43], [44]. In [19], [45], a Frobenius norm is applied to $\mathbf{Z}$, which results in a closed-form solution with a dense coefficient matrix. After obtaining $\mathbf{Z}$, labels of data points can be assigned by a SC based on a constructed affinity matrix $\mathbf{A} = \frac{1}{2}(|\mathbf{Z}| + |\mathbf{Z}|^T)$.

In [7], a GCSC framework is presented to generalize subspace clustering models to the non-Euclidean domain so that more complex relationship can be modeled. The framework can be expressed as

$$\arg\min_{\mathbf{Z}} \|\mathbf{X}\hat{\mathbf{A}}\mathbf{Z} - \mathbf{X}\|_F + \frac{\lambda}{2}\|\mathbf{Z}\|_p,\ \text{s.t. diag}\,(\mathbf{Z}) = 0 \tag{4}$$

where $\hat{\mathbf{A}} = (\mathbf{D}+\mathbf{I})^{-1/2}(\mathbf{A}+\mathbf{I})(\mathbf{D}+\mathbf{I})^{-1/2}$ is an augmented normalized adjacency matrix calculated by the graph adjacent matrix $\mathbf{A}$ and degree matrix $\mathbf{D}$, and λ is a regularization coefficient. The first term in (4) is equivalent to a graph convolutional autoencoder. Noticeably, both $\hat{\mathbf{A}}$ and $\mathbf{Z}$ can be treated as the similarity matrix of data points, however, they have obvious differences. That is, $\hat{\mathbf{A}}$ reveals the intrinsic structure of a graph with self-loops, while $\mathbf{Z}$ signifies the reconstruction contribution of each data in different subspaces (exclude the current data point itself). Furthermore, the use of graph convolution allows us to handle not only nonstructural data but also natural structured data.

III. Hypergraph Convolutional Subspace Clustering (HGCSC)

A. Motivation and Definition

Previous subspace clustering methods seek to find subspaces either in Euclidean space or merely considering simple graph-structured prior. However, raw data or simple graph fail to represent a high-order relationship, e.g., intracluster dependences, thus leading to poor robustness in handling complex HSI datasets. Furthermore, most graph-based subspace clustering models focus mainly on the one-hop neighborhood relationship while ignoring the long-ranging neighboring dependences. These motivate us to generalize subspace clustering into the non-Euclidean domain so that the structural information, especially high-order relationships, can be incorporated into the subspace representations.

The overall framework of the HGCSC approach is shown in Fig. 2. Before elaborating on the details, we first present some necessary notations. Given an HSI, denoted as $\mathbf{X} \in \mathbb{R}^{D\times N}$, where D is the number of spectral bands and N indicates the number of cells, the goal of training is to learn a model $\mathcal{F}_\theta$ to assign each $\boldsymbol{x}_i$ to a certain cluster label y_i, i.e., $\mathcal{F}_\theta : \boldsymbol{x}_i \in \mathbb{R}^D \to y_i \in \{1, 2, \ldots, C\}$, where C is the number of clusters. In the following, we explain hypergraph construction, the definition of hypergraph convolution, hypergraph convolutional self-representation (HGCSR), and the training process.

B. Hypergraph Construction

We first transform an HSI $\mathbf{X} \in \mathbb{R}^{D\times N}$ with D spectral bands and N cells into the irregular domain by constructing a hypergraph $\mathcal{G} = \{\mathbf{X}, \mathbf{H}, \mathbf{W}\}$. Specifically, each cell $\boldsymbol{x}_i$ can be represented as a vertex in a hypergraph and its k nearest neighbors $\mathcal{N}_k(x_i)$ constitute a hyperedge. We adopt Gaussian kernels with the bandwidth parameter σ to measure the similarity of a pair of cells, $s(\boldsymbol{x}_i, \boldsymbol{x}_j)$, i.e.,

$$s\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{\sigma}\right). \tag{5}$$

The hypergraph structure is encoded into an incidence matrix $\mathbf{H} \in \mathbb{R}^{N\times N}$, where $[\mathbf{H}]_{ij} = \{0, 1\}$ denotes whether a cell belongs to a hyperedge, i.e.,

$$[\mathbf{H}]_{ij} = \begin{cases} 1, & \text{if } \boldsymbol{x}_j \in \mathcal{N}_k\left(\boldsymbol{x}_i\right) \\ 0, & \text{otherwise} \end{cases}. \tag{6}$$

Furthermore, we use a sum of similarities of all vertices linked by the same hyperedge to indicate the weight of the hyperedge,

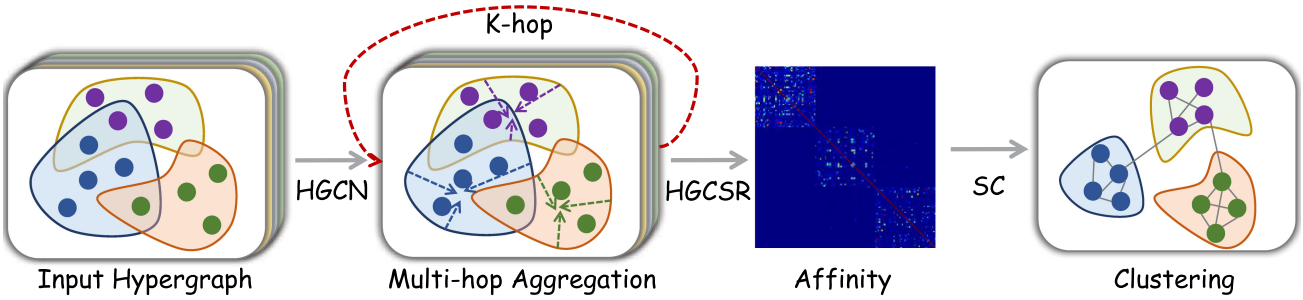


Fig. 2. Overview of the proposed HGCSC approach. In HGCSC, the original HSI data are converted into a hypergraph, where each vertex denotes an HSI cell and each hyperedge links several most similar cells. Subsequently, based on a propagation matrix, a HGCSR with multihop aggregation is adopted to calculate a robust subspace coefficient matrix. Finally, SC is used to assign the cluster labels.

i.e., $[\mathbf{W}]_{ii} = \sum_{\boldsymbol{x}_j \in \mathcal{N}_k(\boldsymbol{x}_i)} s(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Compared with the original HSI, $\mathcal{G}$ contains more high-order information so that we can define the subsequent operations.

C. Definition of Hypergraph Convolution

Following the spectral hypergraph convolution proposed in [36] and [37], we define the nonparametric hypergraph convolution with a symmetric normalization as

$$\hat{\mathbf{X}} = \mathbf{X}\mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_v^{-1/2} \tag{7}$$

where $\hat{\mathbf{X}} \in \mathbb{R}^{D\times N}$ is the embedding of $\mathcal{G}$. For convenience, let $\mathbf{P} = \mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_v^{-1/2}$ be a propagation matrix in $\mathbb{R}^{N\times N}$, then we have $\hat{\mathbf{X}} = \mathbf{X}\mathbf{P}$. The hypergraph convolution can be treated as a two-stage refinement performing "node-hyperedge-node" feature transformation upon the hypergraph structure [46], [47]. The multiplication operation $\mathbf{X}\mathbf{H}$ serves as the information aggregation from nodes to hyperedges, and then, premultiplying $\mathbf{H}^T$ defines to aggregate information from hyperedges to nodes.

D. Hypergraph Convolutional Self-Representation (HGCSR)

We recast the classic self-representation as a structured version by incorporating it with hypergraph convolution. We refer to such a self-representation model as HGCSR. Formally, HGCSR is parameterized by a coefficient matrix $\mathbf{Z} \in \mathbb{R}^{N\times N}$ and is defined as the following reconstruction problem, i.e.,

$$\mathbf{X} = \mathbf{X}\mathbf{P}\mathbf{Z}, \text{ s.t. } \operatorname{diag}(\mathbf{Z}) = 0. \tag{8}$$

Following [19] and [7], by using the Frobenius norm, the $\operatorname{diag}(\mathbf{Z})$ term can be ignored but without loss performance. Then, we can rewrite the HGCSR as a minimum Frobenius norm problem, i.e.,

$$\arg\min_{\mathbf{Z}} \mathcal{J}(\mathbf{Z}) = \frac{1}{2}\|\mathbf{Z}\|_F^2 + \frac{\lambda}{2}\|\mathbf{X} - \mathbf{X}\mathbf{P}\mathbf{Z}\|_F^2. \tag{9}$$

As a result, we can efficiently calculate the closed-form solution for HGCSR. Specifically, the partial derivative with respect to $\mathbf{Z}$ can be denoted as

$$\frac{\partial \mathcal{J}}{\partial \mathbf{Z}} = \left(\mathbf{P}^T\mathbf{X}^T\mathbf{X}\mathbf{P} + \lambda\mathbf{I}_N\right)\mathbf{Z} - \mathbf{P}^T\mathbf{X}^T\mathbf{X}. \tag{10}$$

Thus, we have the closed-form solution, i.e.,

$$\mathbf{Z} = \left(\mathbf{P}^T\mathbf{X}^T\mathbf{X}\mathbf{P} + \lambda\mathbf{I}_N\right)^{-1}\mathbf{P}^T\mathbf{X}^T\mathbf{X}. \tag{11}$$

Algorithm 1: HGCSC.

Input HSI: $\mathbf{X}$, propagation step: K, the number of neighbors: k, and regularization coefficient: λ.
1: Construct hypergraph using (6);
2: Compute propagation matrix $\mathbf{P} = \mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_v^{-1/2}$;
3: Compute multihop propagation matrix $\mathbf{P}^K$;
4: Compute closed-form solution using (13);
5: Construct affinity matrix $\mathbf{A}$;
6: Assign cluster labels using SC;
Output: Clustering results.

E. Multihop Aggregation

In the HGCSR defined previously, we ignore the long-range interdependence between hyperedges and vertices, leading to inefficient utilization. To overcome the problem, following the suggestions in [48] and [49], we provide a simple multihop aggregation rule for HGCSR. More precisely, the strategy stems from a fact that the propagation of a K-layer linear HGCN can be simplified as the K power of the propagation matrix, i.e., $\mathbf{P}^K$. Hence, we use $\mathbf{P}^K$ as the multistep propagation and replace the original single-step propagation in (9)

$$\arg\min_{\mathbf{Z}} \mathcal{J}(\mathbf{Z}) = \frac{1}{2}\|\mathbf{Z}\|_F^2 + \frac{\lambda}{2}\left\|\mathbf{X} - \mathbf{X}\mathbf{P}^K\mathbf{Z}\right\|_F^2. \tag{12}$$

Similarly, its closed-form solution becomes

$$\mathbf{Z} = \left(\left(\mathbf{P}^K\right)^T\mathbf{X}^T\mathbf{X}\mathbf{P}^K + \lambda\mathbf{I}_N\right)^{-1}\left(\mathbf{P}^K\right)^T\mathbf{X}^T\mathbf{X}. \tag{13}$$

It should be noted that the HGCSR with multihop aggregation can be regarded as a simplified deep hypergraph convolutional autoencoder parameterized by learnable parameters $\mathbf{Z}$.

F. Overall Learning Steps

When $\mathbf{Z}$ is obtained, we construct an affinity matrix $\mathbf{A}$ and pass it into the SC. Since there is no globally accepted solution for this step, we adopt the heuristic strategy suggested in [7] and [19] to enhance the block structure. Besides, to fully use spectral-spatial information in HSI, we take a patch of pixels surrounding the central pixel into account. Such a strategy is widely adopted in the HSI classification community [7]. We show the pseudocode of the HGCSC Algorithm 1. It can be seen that the

TABLE I
SUMMARY OF THE FIVE HSI DATASETS USED IN EXPERIMENT

Data sets	SalinasA	Indian Pines	Pavia University	Houston	Xuzhou
Pixels	83×86	85×70	140×150	349×680	100×260
Channels	204	200	103	144	436
Clusters	6	4	8	12	5
Samples	5348	4391	6445	6048	12043
Sensor	AVIRIS	AVIRIS	ROSIS	ITRES CASI-1500	HYSPEX

TABLE II
SETTINGS OF HYPERMETERS IN HGCSC

Data Sets	SalinasA	Indian Pines	Pavia University	Houston	Xuzhou
λ	1000	1000	1000	1000	1000
k	30	20	10	30	10
K	2	2	2	2	2

proposed HCGSC method can be efficiently implemented using sparse matrix multiplication.

G. Remarks and Computational Complexity

HGCSC can be treated as a generalization of traditional subspace clustering models in the irregular domain, particularly on hypergraph. By setting $k = 1$, the propagation matrix $\mathbf{P}$ collapses into an identity matrix, while HGCSC is degraded into the traditional subspace clustering models. Thus, HGCSC is a more general alternative to the traditional subspace clustering with the ability to handle regular data and irregular data. Our HGCSC follows the previous GCSC framework but there exist two main differences. First, HGCSC is defined on hypergraph, while GCSC works on a simple graph. Furthermore, hypergraph is the generalization of the simple graph, signifying that our HGCSC is more general than GCSC. Second, we enable HGCSC to capture long-range relationships, while GCSC only considers 1-hop neighborhoods.

Mathematically, if we solve (13) directly using matrix multiplications, it will lead to a computational complexity upper bounded by $\mathcal{O}(N^3)$. Specifically, the K-hop propagation matrix $\mathbf{P}^K$ needs $\mathcal{O}(KN^2\,M)$ operations. However, its real computational complexity would be significantly reduced by using sparse matrix multiplications and precomputation. Thus, the computational complexity of HGCSC will finally be decreased to be the same as the calculation of a typical least-squares regression, wherein most operations will result from the matrix inverse. Fortunately, there are many well-optimized tools, e.g., SciPy, that can be used to efficiently solve this classic optimization.

IV. EXPERIMENTS

In this section, we extensively evaluate the proposed HGCSC on five frequently used HSI benchmark datasets and compare them with many state-of-the-art approaches. To thoroughly evaluate the performance of each approach, three popular evaluation metrics including overall accuracy (OA), normalized mutual information (NMI), and kappa coefficient (Kappa) are calculated.

A. Setup

1) Datasets: We conduct experiments on five real HSI images, i.e., SalinasA, Indian Pines, Pavia University, Houston2013, and Xuzhou [7], [50]. The first two datasets were acquired by AVIRIS sensors, while the latter three were collected by ROSIS, ITRES CASI-1500, and HYSPEX, respectively. For computational efficiency, we take the subscenes of these datasets as it is done in [7] and [51]–[54]. More details can be found in our previous work [7]. Note that the subscene taken from the Salinas dataset is also known as the SalinasA dataset. The details of the three datasets are summarized in Table I.

In data preprocessing, we perform a PCA to reduce spectral bands into four by preserving at least 95% of the cumulative percentage of variance. With a common solution [55], we use a sliding square window to take samples from an HSI. Specifically, the window size of the Indian Pines is 13×13, while the other four datasets are 9×9. All samples are standardized by scaling feature values into [0,1] by using the min–max scale approach before use.

2) Baselines: We compare the proposed HGCSC model with ten existing clustering methods, including three classical clustering methods and seven popular HSI clustering methods. The classical clustering methods compared in our experiment include SC [14], SSC [16], and LRSC [20]. The popular HSI clustering methods include ℓ_2-norm based SSC (ℓ_2-SSC) [56], unsupervised broad learning (UBL) clustering [52], robust manifold matrix factorization (RMMF) [57], EDSC [19], efficient graph convolutional subspace clustering (EGCSC) [7], deep spatial-spectral subspace clustering (DS^3-Net) [41], and affinity matrix learning via nonnegative matrix factorization (NMFAML) [53]. For these compared methods, we follow their settings reported in the corresponding articles. The hyperparameters of the proposed HGCSC are listed in Table II. All the baseline methods are implemented with Python 3.5 and evaluated on an Intel i5-6500 3.20-GHz CPU with 8-GB RAM.

B. Results

1) Quantitative Results: Table III shows the clustering performance of different methods evaluated on the SalinasA, Indian Pines, Pavia University, Houston, and Xuzhou datasets. It can be observed that, the presented HGCSC model obtains the best clustering performance that significantly outperforms the other clustering methods on all datasets according to OA, NMI, and Kappa values. Furthermore, we can observe the following tendencies.

First, the introduction of hypergraph convolution is beneficial for HSI subspace clustering. Specifically, the proposed HGCSC has significant improvement compared against EGCSC on all datasets. In our experiment, the proposed HGCSC achieves 0.15%, 7.61%, 10.89%, 10.77%, and 13.23% improvement over EGCSC on SalinasA, Indian Pines, Pavia University, Houston, and Xuzhou datasets, respectively. Particularly, HGCSC achieves 100% in OA, NMI, and Kappa value on SalinasA dataset, which is mainly because the land covers in the dataset have distinct subspace structures. This demonstrates that hypergraph convolution is more effective than the simple graph convolution for improving the subspace clustering model.

Second, complementary information and feature extraction of data can remarkably enhance the clustering performance. From all the baselines, we can conclude that by incorporating multiple types of information, clustering methods can obtain a better performance. For example, RMMF and NMFAML tend to achieve higher accuracy than traditional subspace clustering models. It is mainly because that the former combines dimensionality reduction and manifold information, while the latter utilizes homogeneous information. Similarly, the proposed HGCSC is

TABLE III
CLUSTERING PERFORMANCE COMPARISON OF BASELINES AND OUR METHOD ON SALINASA, INDIAN PINES, PAVIA UNIVERSITY, HOUSTON2013, AND XUZHOU DATASETS

Data	Metric	SC	SSC	LRSC	L2-SSC	UBL	RMMF	EDSC	EGCSC	DS^{3}C-Net	NMFAML	HGCSC
SaA.	OA	0.6806	0.7666	0.5613	0.6412	0.9142	0.9820	0.8702	0.9985	0.8698	0.9956	**1**
	NMI	0.7464	0.7571	0.4242	0.6971	0.8692	0.9483	0.9135	0.9949	0.8449	0.9848	**1**
	Kappa	0.6002	0.7138	0.4487	0.5546	0.8943	0.9775	0.8384	0.9981	0.8696	0.9945	**1**
InP.	OA	0.6841	0.4937	0.5142	0.6645	0.6258	0.7121	0.7126	0.8483	0.8388	0.8508	**0.9244**
	NMI	0.5339	0.2261	0.2455	0.3380	0.6680	0.4985	0.4717	0.6442	0.8475	0.7264	**0.8921**
	Kappa	0.5055	0.2913	0.3145	0.5260	0.4690	0.5609	0.5657	0.6422	0.7989	0.7809	**0.8307**
PaU.	OA	0.7691	0.6146	0.4326	0.5842	0.7083	0.7704	0.6175	0.8442	0.8687	0.8967	**0.9531**
	NMI	0.6784	0.6545	0.3793	0.4942	0.6874	0.7388	0.5750	0.8401	0.5682	0.9216	**0.9383**
	Kappa	0.8086	0.4886	0.2549	0.3687	0.6533	0.6804	0.4250	0.7968	0.8685	0.8625	**0.9440**
Hou.	OA	0.3661	0.5526	0.3571	0.4228	NA	0.7039	0.5321	0.6238	NA	0.6346	**0.7315**
	NMI	0.5067	0.7531	0.4472	0.5167	NA	0.7797	0.4772	0.7754	NA	0.7959	0.7004
	Kappa	0.2870	0.5022	0.2851	0.3609	NA	0.6111	0.6768	0.5812	NA	0.5910	**0.8309**
XuZ.	OA	0.7267	0.6353	0.5631	0.5677	NA	0.6111	0.7596	0.7834	NA	0.8178	**0.9157**
	NMI	0.6038	0.4255	0.3631	0.3634	NA	0.6253	0.6834	0.6975	NA	0.7919	**0.8862**
	Kappa	0.5504	0.5226	0.4509	0.4496	NA	0.4361	0.6912	0.6470	NA	0.7459	**0.8470**

[0]The best results are highlighted in bold.

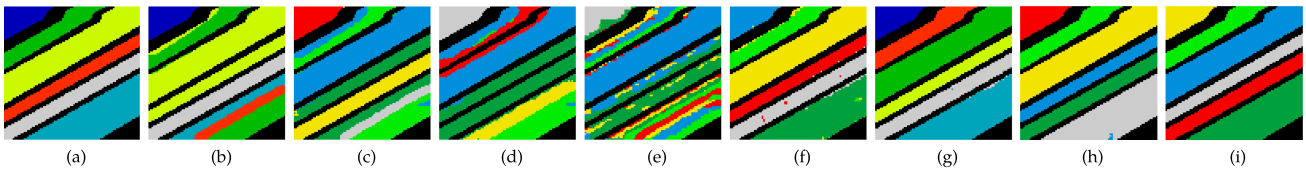


Fig. 3. Clustering results obtained by different methods on the SalinasA dataset. (a) Ground truth. (b) SC 68.06%. (c) SSC 76.66%. (d) ℓ_2-SSC 64.12%. (e) LRSC 56.13%. (f) RMMF 98.20%. (g) NMFAML 99.56%. (h) EGCSC 99.85%. (i) HGCSC 100%.

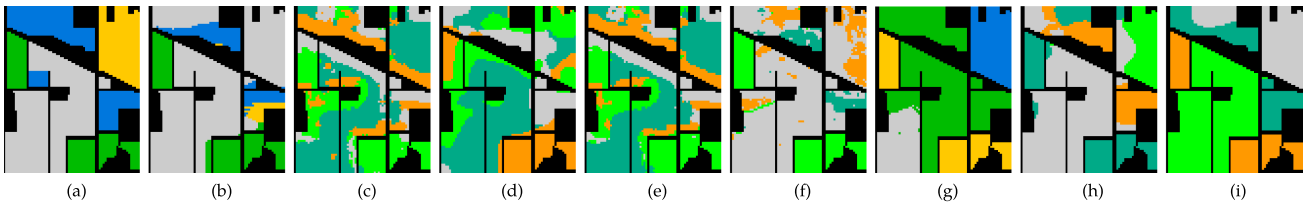


Fig. 4. Clustering results obtained by different methods on the Indian Pines dataset. (a) Ground truth. (b) SC 68.41%. (c) SSC 49.37%. (d) ℓ_2-SSC 66.45%. (e) LRSC 51.42%. (f) RMMF 71.21%. (g) NMFAML 85.08%. (h) EGCSC 84.83%. (i) HGCSC 92.44%.

benefited from the hypergraph structure and our developed information aggregation operations. It shows that an effective information extraction and fusion strategy will be helpful for HSI clustering.

Third, the cluster results obtained by the proposed HGCSC are comparable with a lot of supervised HSI classification methods [1], [58]. To be more specific, the proposed HGCSC achieves 100%, 92.44%, 95.31%, 73.15%, and 91.57% OA values on the SalinasA, Indian Pines, Pavia University, Houston, and Xuzhou datasets, respectively. In general, supervised HSI classification approaches have excellent ability than unsupervised HSI classification approaches because supervised methods use a large number of labeled samples during the training procedure. Nevertheless, our proposed HGCSC provides a new idea for developing an efficient HSI clustering model to compete with supervised HSI classification methods.

2) Qualitative Comparison of Different Methods: In order to further observe the clustering results, we visualize the clustering maps of different clustering methods in Figs. 3–7. Since the source codes of UBL and DS^3-Net have not been released, we cannot reproduce their clustering maps. However, this does not affect the conclusions in this experiment. As we can observe from Fig. 3, the class map obtained by HGCSC on the SalinasA dataset is identical as the ground truth, while the other baseline methods exhibit different levels of noise points. Moreover, the other visual clustering results for the other four datasets are also shown in Figs. 4–7, respectively. For these datasets, the proposed HGCSC shows the best class maps that contain the least noise and are visually the closest to the real ground truths. The class maps obtained by the other clustering methods all show less accurate results compared to HGCSC, and they contain more noise caused by misclustering. On the whole, the hypergraph

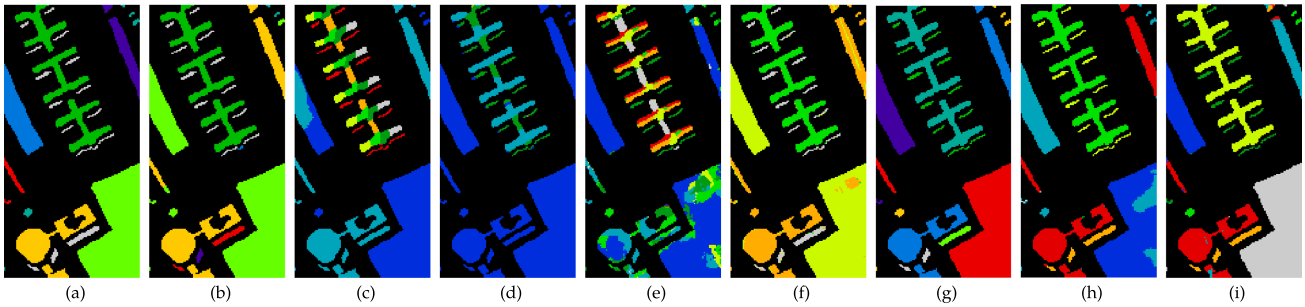


Fig. 5. Clustering results obtained by different methods on the Pavia University dataset. (a) Ground truth. (b) SC 76.91%. (c) SSC 61.46%. (d) ℓ_2-SSC 58.42%. (e) LRSC 43.26%. (f) RMMF 77.04%. (g) NMFAML 89.67%. (h) EGCSC 84.42%. (i) HGCSC 95.31%.

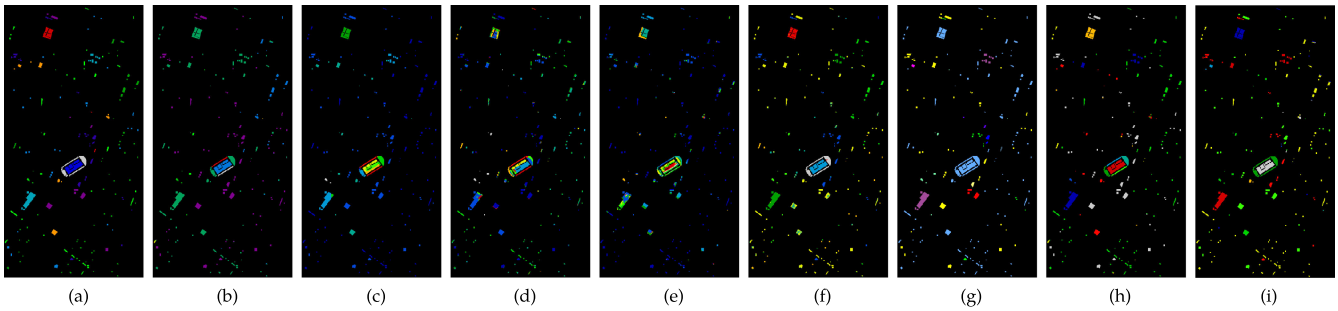


Fig. 6. Clustering results obtained by different methods on the Houston dataset. (a) Ground truth. (b) SC 36.61%. (c) SSC 55.26%. (d) ℓ_2-SSC 42.28%. (e) LRSC 35.71%. (f) RMMF 70.39%. (g) NMFAML 63.46%. (h) EGCSC 62.38%. (i) HGCSC 73.15%.

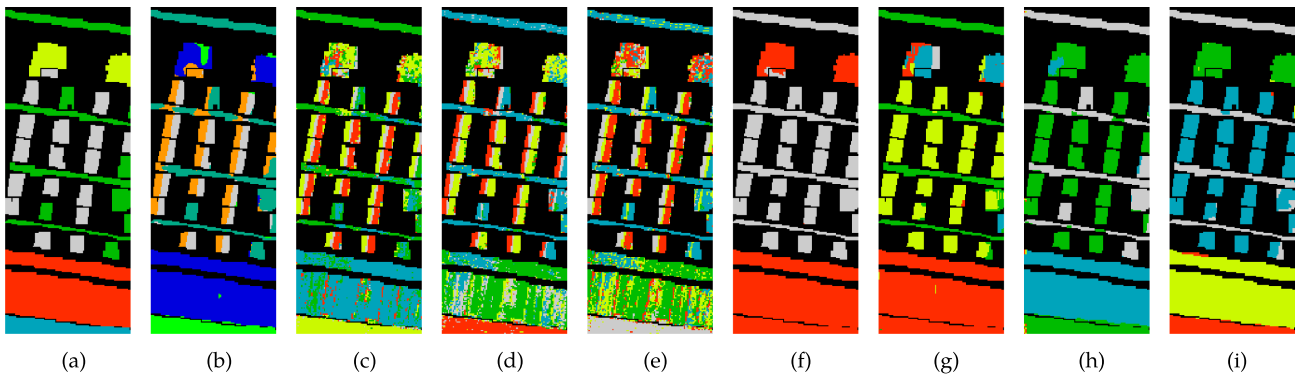


Fig. 7. Clustering results obtained by different methods on the Xuzhou dataset. (a) Ground truth. (b) SC 72.67%. (c) SSC 63.53%. (d) ℓ_2-SSC 56.77%. (e) LRSC 56.31%. (f) RMMF 61.11%. (g) NMFAML 81.78%. (h) EGCSC 78.34%. (i) HGCSC 91.57%.

convolution structure in the proposed HGCSC is beneficial for distinguishing the discriminative spatial and spectral features, which further verifies the effectiveness and superiority of our proposed model for the HSI subspace clustering.

3) Visualization of the Learned Affinity Matrix: To observe the learned affinity matrices, we select two datasets (SalinasA and Houston) and show their affinity matrices obtained by the proposed HGCSC in Fig. 8(a) and (b). Note that all data points are reordered according to the ground truth before calculating the affinity matrix, which makes sense for observing subspace structure. In Fig. 8, each column or row of an affinity matrix represents the contribution using all data points to represent the corresponding data point. Technically, if a group of samples belongs to the same cluster, then the representation coefficients between them are usually nonzero; otherwise, they should be zero. Hence, the ideal affinity matrix is block diagonal. As we can see from Fig. 8(a) and (b), the learned affinity matrices not only are sparse but also have an obviously block-diagonal structure. Furthermore, such subspace structure can be expressed by a hypergraph, and then, be revealed by using the multihop HGCSR. This phenomenon verifies that the proposed HGCSC has the ability to accurately distinguish the internal relationships between clusters, and thus, performs superior subspace clustering results.

TABLE IV
ROBUSTNESS COMPARISON WITH DIFFERENT INTENSITY NOISE ON INDIAN PINES DATASET

Method	$variance = 0$			$variance = 0.1$			$variance = 0.3$			$variance = 0.5$		
	OA	NMI	Kappa	OA	NMI	Kappa	OA	NMI	Kappa	OA	NMI	Kappa
HGCSC	0.9244	0.8921	0.8307	0.8242	0.7471	0.5848	0.7606	0.6604	0.4819	0.7053	0.5663	0.3510
EGCSC	0.8483	0.6442	0.6422	0.6773	0.5754	0.4965	0.6702	0.5042	0.3733	0.655	0.4795	0.0318
EDSC	0.7126	0.4717	0.5657	0.6106	0.3835	0.4552	0.4348	0.0178	0.3268	0.4439	0.0321	0.0451

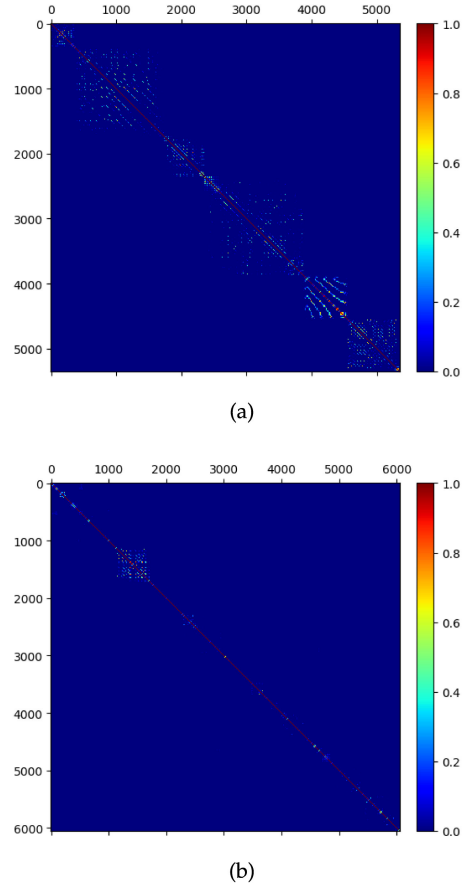


Fig. 8. Visualization of the affinity matrix learned on (a) SalinasA. (b) Houston datsets.

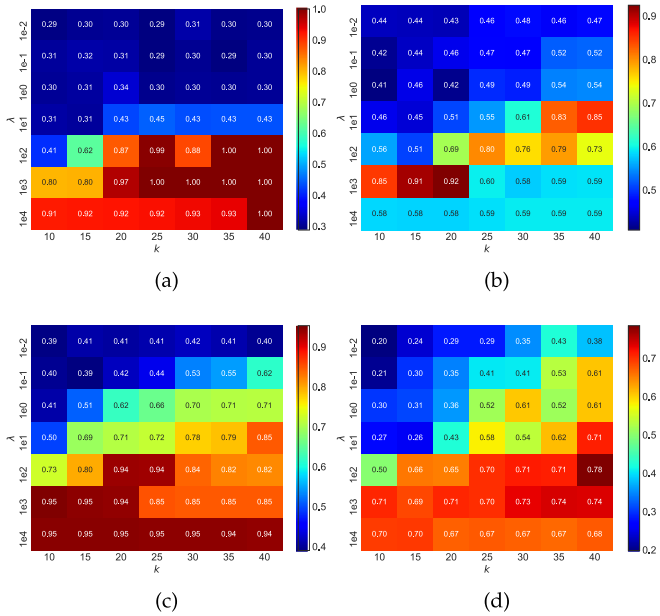


Fig. 9. Influence of hyperparameters of k and λ for HGCSC on (a) SalinasA, (b) Indian Pines, (c) Pavia University, and (d) Houston datasets.

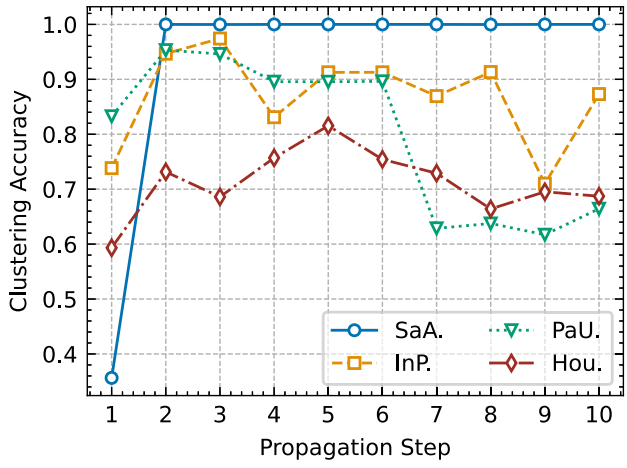


Fig. 10. Influence of the propagation steps on SalinasA, Indian Pines, Pavia University, and Houston datasets.

C. Ablation Studies

1) Impact of λ and k: Fig. 9(a)–(d) exhibits the impact of the two important hyperparameters of HGCSC, i.e., the regularization coefficient λ and the number of the nearest neighbors k. In this experiment, we set $\lambda = [10^{-2}, 10^{-1}, \ldots, 10^{4}]$ and $k = [10, 15, \ldots, 40]$ for all HSI datasets. As shown in Fig. 9, we can see that the regularization coefficient λ has a significant effect on clustering performance. In general, the clustering accuracy of HGCSC tends to increase as λ increased. Taken overall, the performance of HGCSC degrades dramatically when $\lambda < 10$. Therefore, empirically, it is appropriate to set λ within $[10^2, 10^4]$. In addition, the neighborhood size k reveals the property of the predefined hypergraph, thus, it will impact the clustering accuracy. If the number of cells belonging to a certain class is less than k, then the predefined hypergraph will inevitably contain noisy hyperedges, resulting in an unreliable hypergraph structure. Consequently, the neighborhood size should not be too large. According to the empirical study, the suitable k can be found in a range from 10 to 30.

2) Impact of Propagation Step: The propagation operation that aggregates information from high order neighborhoods suffers from performance drops when the propagation step is too large. We thus test the robustness of the HGCSC model on large

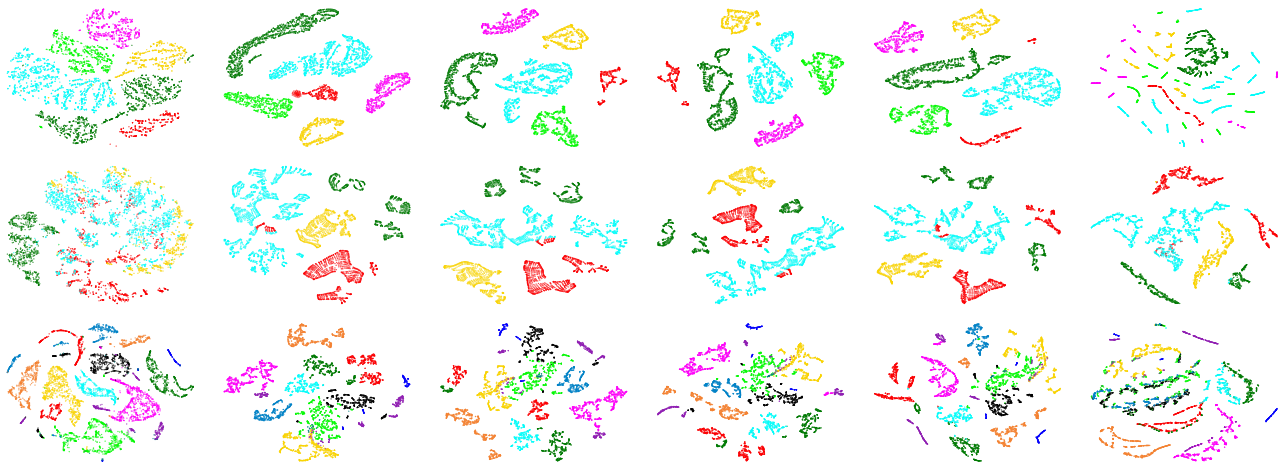


Fig. 11. t-SNE visualization of the embeddings obtained on the SalinasA (the first row), Indian Pines (the second row), and Houston (the third row) with an increasing propagation steps. From left to right, the embeddings are obtained under $K = 0$, 1, 2, 3, 5, and 10, orderly. Each color indicates a cluster.

propagation steps. In Fig. 10, we show the clustering accuracy curves with different propagation steps varying from 0 to 10. We find that, like most GCN models [28], our model performance is remarkably improved by small propagation steps (e.g., $K \leq 3$), while suffers from performance drops with large propagation steps. Such phenomenon is caused by the oversmoothing problem commonly encountered in graph convolution, which often leads to similar and indistinguishable node representation for different clusters. Instead, when a suitable propagation step is set, e.g., $K = 2$, the long-range interdependence is incorporated into the self-representation, and resulting in better robustness. Notably, when $K = 0$, HGCSC degenerates to the classical SSC model while its performance is significantly worse than that with hypergraph convolution. It signifies that our HGCSC is superior to the classic subspace clustering models.

To better understand the effect of propagation steps on the clustering, we visualize the embeddings obtained under different steps. From Fig. 11, we can observe that the proposed multihop aggregation can distinctly increase the separability between different clusters. More precisely, the initial features ($K = 0$) can be roughly separated into a few clusters. However, there are also many samples of different clusters mixed with others. Compared with the initial features, hypergraph embeddings show better intraclass compactness and interclass separability, which is beneficial to the subsequent subspace clustering. Furthermore, such a tendency is enhanced as the propagation steps increase. This shows the effectiveness of the multihop aggregation strategy. Nonetheless, from the last column in figures ($K = 10$), we can see an obvious oversmoothing phenomenon, especially on SalinasA and Houston datasets.

3) Analysis of Robustness: To explore the robustness of HGCSC, we conduct clustering on the Indian Pines dataset with different noise levels. In this experiment, we add random Gaussian noises with zero means and variances of 0.1, 0.3, and 0.5 into the dataset. Higher variance means that the data contain more noise. From Table IV, we can see that HGCSC consistently shows better robustness to different intensity noise than EDSC and EGCSC. This is because HGCSC leverages both high-order information and long-range dependences, which are more robust than the simple graph and pure features. This shows the effectiveness of our method. Furthermore, EGCSC remarkably improves the robustness over EDSC by using the simple graph structure, demonstrating the importance of structural information.

V. Conclusion

In this article, we proposed a simple yet very effective HSI clustering approach, namely HGCSC, which learns the subspace representation coefficients by using a multihop HGCSC model. Experimental results on five popular HSI datasets show that our HGCSC not only can effectively enhance a strong model, i.e., GCSC but also achieve state-of-the-art performances that consistently outperform competitors by a large margin. Due to the limitation of computational resources, we only tested HGCSC on small-scale scenes in this work. We will try to scale it to large-scale datasets in the future. In this work, we focused on the assumption that data points are drawn from a linear subspace. We will apply HGCSC to nonlinear graph neural network architectures.

References

[1] P. Ghamisi *et al.*, "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.

[2] S. Salcedo-Sanz *et al.*, "Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources," *Inf. Fusion*, vol. 63, pp. 256–272, 2020.

[3] S. Selvakumaran *et al.*, "Combined InSAR and terrestrial structural monitoring of bridges," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7141–7153, Oct. 2020.

[4] S. Saha, F. Bovolo, and L. Bruzzone, "Building change detection in VHR SAR images via unsupervised deep transcoding," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 1917–1929, Mar. 2021.

[5] Q. Wang *et al.*, "Identification of melanoma from hyperspectral pathology image using 3D convolutional networks," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 218–227, Jan. 2021.

[6] M. Zhang, M. Gong, H. He, and S. Zhu, "Symmetric all convolutional neural-network-based unsupervised feature extraction for hyperspectral images classification," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2020.3020540.

[7] 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, doi: 10.1109/TGRS.2020.3018135.

[8] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[9] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[10] H. Zhai, H. Zhang, P. LI, and L. Zhang, "Hyperspectral image clustering: Current achievements and future lines," *IEEE Geosci. Remote Sens. Mag.*, to be published, doi: 10.1109/MGRS.2020.3032575.

[11] Y. Cai, Z. Zhang, Z. Cai, X. Liu, and X. Jiang, "Hypergraph-structured autoencoder for unsupervised and semisupervised classification of hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2021.3054868.

[12] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.

[13] S. Ghaffarian and S. Ghaffarian, "Automatic histogram-based fuzzy C-means clustering for remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 97, pp. 46–57, 2014.

[14] 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.

[15] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Mar. 2011.

[16] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[17] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Joint and progressive subspace analysis (JPSA) with spatial-spectral manifold alignment for semisupervised hyperspectral dimensionality reduction," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3602–3615, Jul. 2021.

[18] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.

[19] P. Ji, M.Salzmann, and H. Li, "Efficient dense subspace clustering," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2014, pp. 461–468.

[20] R. Vidal and P. Favaro, "Low rank subspace clustering (LRSC)," *Pattern Recognit. Lett.*, vol. 43, pp. 47–61, 2014.

[21] C. Lu, J. Feng, Z. Lin, T. Mei, and S. Yan, "Subspace clustering by block diagonal representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 487–501, Feb. 2019.

[22] S. Huang, H. Zhang, and A. Pižurica, "Joint sparsity based sparse subspace clustering for hyperspectral images," in *Proc. 25th IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 3878–3882.

[23] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Nonlocal means regularized sketched reweighted sparse and low-rank subspace clustering for large hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4164–4178, May 2020.

[24] P. Ji, T. Zhang, H. M. Li Salzmann, and I. Reid, "Deep subspace clustering networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 24–33.

[25] Y. Cai, M. Zeng, Z. Cai, X. Liu, and Z. Zhang, "Graph regularized residual subspace clustering network for hyperspectral image clustering," *Inf. Sci.*, vol. 578, pp. 85–101, 2021.

[26] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 504–517, Mar. 2016.

[27] X. Jiang, X. Song, Y. Zhang, J. Jiang, J. Gao, and Z. Cai, "Laplacian regularized spatial-aware collaborative graph for discriminant analysis of hyperspectral imagery," *Remote Sens.*, vol. 11, no. 1, 2019.

[28] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learn. Representations*, Toulon, France, pp. 1–10, 2017.

[29] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[30] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[31] L. Mou, X. Lu, X. Li, and X. X. Zhu, "Nonlocal graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8246–8257, Dec. 2020.

[32] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Multiscale graph sample and aggregate network with context-aware learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4561–4572, Apr. 2021.

[33] T. Jin, L. Cao, B. Zhang, X. Sun, C. Deng, and R. Ji, "Hypergraph induced convolutional manifold networks," in *Proc. Int. Joint Conf. Artif. Intell.*, 2019, pp. 2670–2676.

[34] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, vol. 19, pp. 1601–1608.

[35] S. Huang, H. Zhang, and A. Pižurica, "Hybrid-hypergraph regularized multiview subspace clustering for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3074184.

[36] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 01, pp. 3558–3565.

[37] S. Bai, F. Zhang, and P. H. Torr, "Hypergraph convolution and hypergraph attention," *Pattern Recognit.*, vol. 110, 2021, Art. no. 107637.

[38] H. Zhai, H. Zhang, X. Xu, L. Zhang, and P. Li, "Kernel sparse subspace clustering with a spatial max pooling operation for hyperspectral remote sensing data interpretation," *Remote Sens.*, vol. 9, no. 4, 2017, Art. no. 335.

[39] Y. Wan, Y. Zhong, A. Ma, and L. Zhang, "Multi-objective sparse subspace clustering for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2290–2307, Apr. 2020.

[40] X. Zhu, S. Zhang, Y. Li, J. Zhang, L. Yang, and Y. Fang, "Low-rank sparse subspace for spectral clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 8, pp. 1532–1543, Aug. 2019.

[41] J. Lei, X. Li, B. Peng, L. Fang, N. Ling, and Q. Huang, "Deep spatial-spectral subspace clustering for hyperspectral image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 7, pp. 2686–2697, Jul. 2021.

[42] 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, Mar. 2021.

[43] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 773–777, May 2017.

[44] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 544–557, Jan. 2016.

[45] Y. Cai, Z. Zhang, X. Liu, and Z. Cai, "Efficient graph convolutional self-representation for band selection of hyperspectral image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4869–4880, Aug. 2020, doi: 10.1109/JSTARS.2020.3018229.

[46] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, and X. Zhang, "Self-supervised hypergraph convolutional networks for session-based recommendation," in *Proc. 35th AAAI Conf. Artif. Intell. (AAAI-21)*, 2021, pp. 4503–4511.

[47] J. Yu, H. Yin, J. Li, Q. Wang, N. Q. V. Hung, and X. Zhang, *Self-Supervised Multi-Channel Hypergraph Convolutional Network for Social Recommendation*. New York, NY, USA: Assoc. Comput. Mach., 2021, pp. 413–424.

[48] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6861–6871.

[49] J. Klicpera, A. Bojchevski, and S. Günnemann, "Predict then propagate: Graph neural networks meet personalized pagerank," in *Proc. Int. Conf. Learn. Representations*, 2019, pp. 1–11.

[50] K. Tan, F. Wu, Q. Du, P. Du, and Y. Chen, "A parallel Gaussian-Bernoulli restricted Boltzmann machine for mining area classification with hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 627–636, Feb. 2019.

[51] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1723–1740, Mar. 2019.

[52] Y. Kong, Y. Cheng, C. L. P. Chen, and X. Wang, "Hyperspectral image clustering based on unsupervised broad learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1741–1745, Nov. 2019.

[53] Y. Qin, B. Li, W. Ni, S. Quan, P. Wang, and H. Bian, "Affinity matrix learning via nonnegative matrix factorization for hyperspectral imagery clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 402–415, Nov. 2021.

[54] Y. Pan, Y. Jiao, T. Li, and Y. Gu, "An efficient algorithm for hyperspectral image clustering," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2019, pp. 2167–2171.
[55] 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.
[56] 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.
[57] 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, 2019.
[58] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.

Zijia Zhang received the M.S. degrees in electronic and communication engineering from the Hubei University, Wuhan, China, in 2018. She is currently working toward the Ph.D. degree in computer science with the School of Computer Science, China University of Geosciences, Wuhan, China.

From August 2021 to August 2022, she was a Visiting Ph.D. Student with Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Freiberg, Germany. Her research interests include evolutionary computation, machine learning, and hyperspectral image processing.

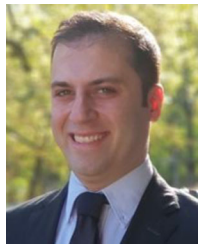
Yaoming Cai (Student Member, IEEE) received the B.Eng. degree in information security in 2016, from the China University of Geosciences, Wuhan, China, where he is currently working toward the Ph.D. degree in computer science with the School of Computer Science.

From August 2021 to August 2022, he was a Visiting Ph.D. Student with Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Freiberg, Germany. His research interests include machine learning, pattern recognition, evolutionary computation, and hyperspectral image processing.

Wenyin Gong (Member, IEEE) received the B.Eng., M.Eng., and Ph.D. degrees in computer science from the China University of Geosciences, Wuhan, China, in 2004, 2007, and 2010, respectively.

He is currently a Professor with School of Computer Science, China University of Geosciences. His research interests include evolutionary algorithms, evolutionary optimization, and their applications. He has authored and co-authored more than 80 research papers in journals and international conferences.

Dr. Gong served as a Referee for more than 40 international journals, such as IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE, *ACM Transactions on Intelligent Systems and Technology*, *Information Sciences*, *European Journal of Operational Research*, *Applied Soft Computing*, *Journal of Power Sources*, etc.



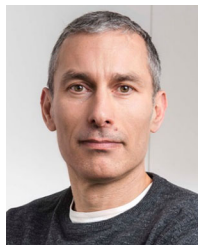
Pedram Ghamisi (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015.

He works as the Head of the Machine Learning Group, Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Freiberg, Germany and as a Visiting Professor and Group Leader with AI4RS, Institute of Advanced Research in Artificial Intelligence, Vienna, Austria. He is a Co-Founder of VasoGnosis Inc., with two branches in San Jose and Milwaukee, USA. His research interests include interdisciplinary research on machine (deep) learning, image and signal processing, and multisensor data fusion.

Dr. Ghamisi was the Co-Chair of the IEEE Image Analysis and Data Fusion Committee (IEEE IADF) between 2019 and 2021. He was the recipient of the IEEE Mikio Takagi Prize for winning the Student Paper Competition at the IEEE International Geoscience and Remote Sensing Symposium in 2013, the first prize of the data fusion contest organized by the IEEE IADF in 2017, the Best Reviewer Prize of IEEE Geoscience and Remote Sensing Letters in 2017, and the IEEE Geoscience and Remote Sensing Society 2020 Highest-Impact Paper Award. He is an Associate Editor for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Xiaobo Liu (Senior Member, IEEE) received the M.S. degree in computer science and the Ph.D. degree in geosciences information engineering from the China University of Geosciences, Wuhan, China, in 2008 and 2012, respectively.

He is currently an Associate Professor with the School of Automation, China University of Geosciences. His research interests include machine learning, evolutionary computation, and hyperspectral remote sensing image processing.



Richard Gloaguen received the Ph.D. degree (Doctor Communitatis Europae) in marine geosciences from the University of Western Brittany, Brest, France, in collaboration with the Royal Holloway University of London, Egham, U.K., and Göttingen University, Göttingen, Germany, in 2000.

He was a Marie Curie Post-Doctoral Research Associate with the Royal Holloway University of London from 2000 to 2003. He led the Remote Sensing Group, University Bergakademie Freiberg, Freiberg, Germany, from 2003 to 2013. Since 2013, he has been leading the division "Exploration Technology," Helmholtz- Institute Freiberg for Resource Technology, Freiberg. He is currently involved in unmanned-aerial-vehicle-based multisource imaging, laser-induced fluorescence, and noninvasive exploration. His research interests include multisource and multiscale remote sensing integration.