

Graph-Based Active Learning for Nearly Blind Hyperspectral Unmixing

Bohan Chen¹, Student Member, IEEE, Yifei Lou², Andrea L. Bertozzi³, Member, IEEE,
and Jocelyn Chanussot⁴, Fellow, IEEE

Abstract—Hyperspectral unmixing (HSU) is an effective tool to ascertain the material composition of each pixel in a hyperspectral image with typically hundreds of spectral channels. In this article, we propose two graph-based semisupervised unmixing methods. The first one directly applies graph learning to the unmixing problem, while the second one solves an optimization problem that combines the linear unmixing model and a graph-based regularization term. Following a semisupervised framework, our methods require a very small number of training pixels that can be selected by a graph-based active learning method. We assume to obtain the ground-truth information at these selected pixels, which can be either the exact (EXT) abundance value or the one-hot (OH) pseudo-label. In practice, the latter is much easier to obtain, which can be achieved by minimally involving a human in the loop. Compared with other popular blind unmixing methods, our methods significantly improve performance with minimal supervision. Specifically, the experiments demonstrate that the proposed methods improve the state-of-the-art blind unmixing approaches by 50% or more using only 0.4% of training pixels.

Index Terms—Active learning, graph learning, hyperspectral unmixing (HSU), semisupervised learning.

I. INTRODUCTION

DATA obtained by hyperspectral sensors provide both spatial and spectral representations of a scene. Compared with regular color images, which only have three color channels (red, green, and blue), hyperspectral images often contain hundreds to thousands of spectral channels. However, hyperspectral imaging is limited by its low spatial resolution, and hence, hyperspectral unmixing (HSU) is an effective tool

to identify the pure materials and estimate the proportions of constituent end-members at each pixel, also known as the *abundance map*. The spectral signature of a pure material is called *end-member*, which can often be measured under a laboratory setting. Unfortunately, the ground-truth end-member is often unavailable due to its large variability in any real scenario. The blind unmixing process involves the estimation of all the end-members and the abundance map simultaneously.

A. Literature Review of HSU

In our study, we employ a *linear* mixing model for unmixing, wherein each pixel's spectral measurement is represented as a linear combination of constituent end-members. Given the physical interpretation of hyperspectral mixing, we impose nonnegativity constraints on both end-members and the abundance map. In addition, we apply a sum-to-one constraint, a common practice in HSU, signifying that each pixel's abundance vector resides within the probability simplex. There are some extended linear mixing models that consider end-memberwise scaling factors [1] and the spectral variability [2]. Note that these nonlinear mixing models [3], [4] rely on more complicated assumptions about how light rays interact with end-members. It is also plausible to remove the sum-to-one constraint when illumination conditions or the topography of the scene changes locally [1].

Specifically for blind HSU, it is natural to apply the nonnegative matrix factorization (NMF) [5], [6] that decomposes the data matrix into a product of two matrices with nonnegative entries (one encodes the end-member matrix, and the other is the abundance map) [7], [8], [9]. However, even with the nonnegativity and sum-to-one constraints, blind HSU is a highly ill-posed inverse problem, and hence, a variety of regularizations have been proposed to refine the solution space. One classic method is the ℓ_2 norm in fully constrained least squares unmixing (FCLS) [10]. Furthermore, spatial sparsity of abundances is a reasonable assumption due to the fact that only a few end-members could appear in a single pixel. Some popular sparsity-promoting regularizations used in HSU include the ℓ_0 norm [11], the ℓ_1 norm [12], the $\ell_{1/2}$ norm [13], and the mixed $\ell_{p,q}$ norm for group sparsity [14]. By treating the abundance map for each material as an image, total variation (TV) regularization [15] has been applied to HSU for spatial continuity and edge preservation. TV-related approaches include sparse unmixing via variable

Manuscript received 20 April 2023; revised 28 August 2023; accepted 2 September 2023. Date of publication 11 September 2023; date of current version 26 September 2023. This work was supported in part by MIAI@Grenoble Alpes under Grant ANR-19-P3IA-0003. The work of Bohan Chen was supported by the UC-National Laboratory In-Residence Graduate Fellowship under Grant L21GF3606. The work of Yifei Lou was supported in part by NSF CAREER under Award 1846690. The work of Andrea L. Bertozzi was supported by an academic grant from the National Geospatial-Intelligence Agency through the Project "Active Learning Methods on Graphs for Image, Video and Multispectral Datasets" under Award HM0476-21-1-0003. Approved for public release on May 12, 2023, NGA-U-2023-01028. (Corresponding author: Bohan Chen.)

Bohan Chen and Andrea L. Bertozzi are with the Department of Mathematics, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 USA (e-mail: bhchenyz@ucla.edu).

Yifei Lou is with the Department of Mathematics and the School of Data Science and Society, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

Jocelyn Chanussot is with CNRS, Grenoble Institute of Technology (Grenoble INP), GIPSA-Lab, Université Grenoble Alpes, 38000 Grenoble, France.

Digital Object Identifier 10.1109/TGRS.2023.3313933

splitting augmented Lagrangian and TV (SUnSAL-TV) [16], TV with sparse NMF [17], TV with nonnegative tensor factorization [18], and an improved collaborative NMF with TV (ICoNMF-TV) [19]. Recently, TV is reformulated as a quadratic regularization that promotes the minimum volume in the NMF framework, referred to as quadratic minimum volume (QMV) [20].

Graph-based approaches [21] also play an important role in HSU. TV has been extended from vectors in Euclidean space to signals defined on a graph. For example, the graph TV (gTV) [22] is a special case of the p -Dirichlet form [23], [24] in graph signal processing. Some graph regularization techniques for hyperspectral imaging include structured sparse-regularized NMF (SS-NMF) [25] and graph-regularized $\ell_{1/2}$ NMF (GLNMF) [26]. Graph-based approaches, while powerful, can suffer from intensive computation, particularly when computing pairwise similarity between pixels. Strategies in speeding up the weight computation include the use of superpixels [27] rather than using the entire hyperspectral image and the Nyström method [28] to generate low-rank approximations of the graph Laplacian [29], [30]. Another efficient alternative is the use of sparse weight matrices, such as the K -nearest neighbors (KNNs) weight matrix [31], [32].

In recent years, neural networks have been applied to the blind unmixing problem, such as two-staged self-supervised networks [33], [34], a minimal simplex convolutional neural network [35], a two-stream Siamese deep network [36], and attention networks [37]. Furthermore, some semisupervised advanced deep learning methods [38], [39] have integrated the use of the graph Laplacian and exhibited remarkable potential in HSU. None of these methods address the issues discussed in the next paragraph.

B. Motivation and Our Contributions

There are limitations of the existing blind HSU methods in many real-world scenarios. For example, unmixing often requires estimation of the number of end-members [40], while the abundance maps require human experts to convey meaning to each end-member. Meanwhile, acquiring the ground-truth abundance maps or end-member spectra is a great challenge [41], [42], making it extremely difficult to train fully supervised models. Such drawback motivates us to consider a semisupervised model with pseudo-labels. These are representative pixels for each end-member, which can be easily obtained by expert's visual inspection of the data. Prior knowledge of the number of end-members is naturally included in the pseudo-labels. We further adopt an active learning [43], [44], [45] approach to reinforce semisupervised machine learning methods by carefully automating the selection its training set. Active learning has been successfully applied to hyperspectral image classification and segmentation tasks [46], [47], [48], [49], [50]. The core of active learning is to sample data points according to an acquisition function [51], [52] that automates the introduction of new training data during the algorithm. We believe active learning and pseudo-labels would be an ideal combination to improve the performance of HSU.

The key problem we solve in this article is to maximize the improvement of our estimated abundance maps and end-member spectra using minimal supervision. With the training pixels selected by active learning, we propose two semisupervised HSU models. We refer to these methods as *nearly blind HSU*, since both of them require a very small number of training labels, and accept either pseudo-one-hot (OH) labels or ground-truth abundance maps. Our first model, called graph learning unmixing (GLU), takes the output of a graph learning method [21] directly as the abundance maps for the HSU, followed by the estimation of the end-member matrix. Our second model, called graph-regularized semisupervised unmixing (GRSU), combines the linear unmixing model, a graph-based regularization term, and a loss function applied to the label set (obtained by active learning) by solving a joint optimization problem. The flowchart of our proposed models (GLU and GRSU) is illustrated in Fig. 1. We summarize the novelties of this work as follows.

- 1) We present an effective pipeline (Section III-A) to select labeled pixels by graph-based active learning for the HSU problem.
- 2) We apply the idea of graph Laplace learning to the HSU problem by taking its output (class probability) as the estimated abundance map. Based on this idea, we develop the GLU model (Section III-B).
- 3) We develop a novel semisupervised HSU model, GRSU (Section III-C), by combining graph-based regularization terms with the linear mixing model and a small number of labeled pixels.
- 4) Our proposed methods, GLU and GRSU, bear significant practical implications. By utilizing only a small number of easily obtainable pseudo-labels, our methods markedly improve the HSU performance.

The rest of this article is organized as follows. Section II reviews knowledge about graph learning and active learning. Section III introduces our regularized graph Laplace learning together with semisupervised unmixing algorithms. Section IV shows the experimental results that compare our proposed algorithms and other blind HSU methods. Finally, Section V concludes this article.

II. BACKGROUND AND NOTATION FOR GRAPH ACTIVE LEARNING

This section reviews the related works on graph construction, graph learning, and graph-based active learning approaches. These techniques are key ingredients in our proposed semisupervised unmixing models (detailed in Section III).

A. Graph Construction Using KNN Similarity Weight

Given a set of N nodes or vertices $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ with $\mathbf{x}_i \in \mathbb{R}^{p \times 1}$ for $i = 1, \dots, N$, we construct a graph $G = (\mathcal{X}, W)$, in which the similarity weight matrix W is computed by KNNs. Specifically, we define the angular distance function $d(\mathbf{x}_i, \mathbf{x}_j)$ as the angle between \mathbf{x}_i and \mathbf{x}_j , i.e.,

$$d(\mathbf{x}_i, \mathbf{x}_j) = \arccos\left(\frac{\mathbf{x}_i^\top \mathbf{x}_j}{\|\mathbf{x}_i\|_2 \|\mathbf{x}_j\|_2}\right). \quad (1)$$

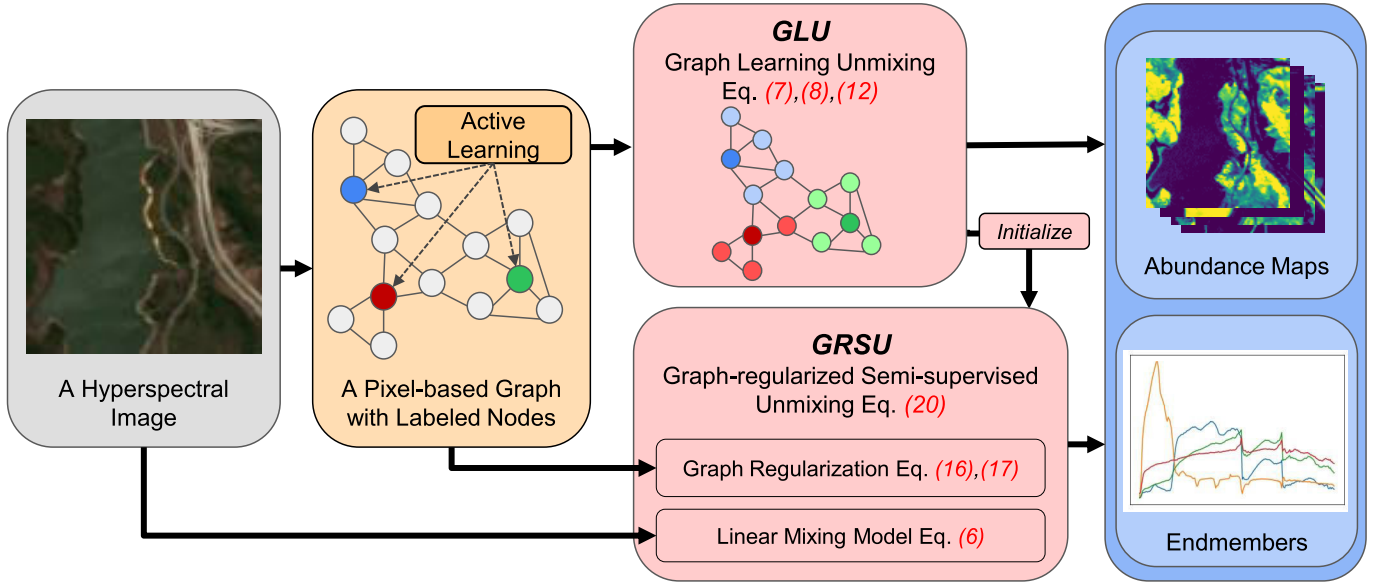


Fig. 1. Flowchart of our semisupervised HSU models. The gray box indicates an input hyperspectral image. The orange boxes are the graph construction and graph-based active learning to select labeled nodes (pixels) for the training process (Sections II-A and III-A). Two red boxes are our proposed models, GLU (Section III-B) and GRSU (Section III-C). GLU applies graph Laplace learning directly to the unmixing task, while GRSU combines the graph-based regularization term with the linear unmixing model from hyperspectral imaging into a joint optimization to be solved. GLU also serves as the initialization of the GRSU optimization process. The blue boxes are the outputs of GLU and GRSU, i.e., estimated end-members and abundance map.

We only consider edges between \mathbf{x}_i and its KNNs according to the angular distance $d(\mathbf{x}_i, \cdot)$, which can be implemented by an approximate nearest neighbor search algorithm [31]. For each node \mathbf{x}_i , we denote $\{\mathbf{x}_{i_k}\}$, $k = 1, 2, \dots, K$, as the KNNs of \mathbf{x}_i (including \mathbf{x}_i itself). We define a weight value W_{ij}^{KNN} between nodes \mathbf{x}_i and \mathbf{x}_j as follows:

$$W_{ij}^{\text{KNN}} = \begin{cases} \exp\left(-\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{\sigma_i \sigma_j}\right), & j = i_1, i_2, \dots, i_K \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $\sigma_i = (d(\mathbf{x}_i, \mathbf{x}_{i_K}))^{1/2}$. We define a new weight by $W_{ij} := (W_{ij}^{\text{KNN}} + W_{ji}^{\text{KNN}})/2$, so that the weight matrix W is symmetric. Note that W is sparse and nonnegative (i.e., $W_{ij} \geq 0$).

Given the weight matrix W , the degree matrix D is a diagonal matrix, whose diagonal entry is $d_i = \sum_{j=1}^N W_{ij}$ for $i = 1, \dots, N$. We further define the graph Laplacian matrix [53], [54] by $L = D - W$.

B. Graph Laplace Learning

With the graph $G = (\mathcal{X}, W)$ constructed in Section II-A, we review a graph-based approach for semisupervised learning, referred to as graph Laplace learning [55]. We suppose ground-truth labels of q classes are available on a subset of vertices $\hat{\mathcal{X}} \subset \mathcal{X}$. For every $\hat{\mathbf{x}}_j \in \hat{\mathcal{X}}$, its label is denoted by $y_j^\dagger \in \{1, 2, \dots, q\}$, and its corresponding OH vector is defined as $\mathbf{y}_j^\dagger = \mathbf{e}_{y_j^\dagger}$, where \mathbf{e}_k is the k th standard basis vector with all zeros except a 1 at the i th entry. The goal for graph learning is to predict the labels of the *unlabeled* set $\mathcal{X} \setminus \hat{\mathcal{X}}$.

The classification of nodes is based on a node function $\mathbf{u} : \mathcal{X} \rightarrow \mathbb{R}^q$. Then, the predicted label of $\mathbf{x}_i \in \mathcal{X}$ is $y_i = \arg \max_k \{\mathbf{u}_k(\mathbf{x}_i), k = 1, 2, \dots, q\}$, where $\mathbf{u}_k(\cdot)$ is the k th entry of $\mathbf{u}(\mathbf{x}_i)$. The graph Laplace learning model [55] obtains

the optimal node function \mathbf{u}^* by minimizing the following objective function:

$$J_\ell(\mathbf{u}; \mathbf{y}^\dagger) = \frac{1}{4} \sum_{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}} W_{ij} \|\mathbf{u}(\mathbf{x}_i) - \mathbf{u}(\mathbf{x}_j)\|_2^2 + \sum_{\hat{\mathbf{x}}_j \in \hat{\mathcal{X}}} \ell(\mathbf{u}(\hat{\mathbf{x}}_j), \mathbf{y}_j^\dagger). \quad (3)$$

The first term of $J_\ell(\mathbf{u}; \mathbf{y}^\dagger)$ in (3) quantifies the smoothness on the node function in the sense that $\mathbf{u}(\mathbf{x}_i)$ is closer to $\mathbf{u}(\mathbf{x}_j)$ for a larger weight W_{ij} between the two nodes. The second term of $J_\ell(\mathbf{u}; \mathbf{y}^\dagger)$ uses the loss function ℓ to measure the difference between the prediction $\mathbf{u}(\hat{\mathbf{x}}_j)$ and the ground truth \mathbf{y}_j^\dagger for $\hat{\mathbf{x}}_j \in \hat{\mathcal{X}}$. We follow [55] to choose the hard-constraint penalty as the loss function, that is:

$$\ell(\mathbf{x}, \mathbf{y}) = \begin{cases} +\infty, & \text{if } \mathbf{x} \neq \mathbf{y} \\ 0, & \text{if } \mathbf{x} = \mathbf{y}. \end{cases} \quad (4)$$

The penalty function ℓ_h forces the minimizer to take $\mathbf{u}^*(\hat{\mathbf{x}}_j) = \mathbf{y}_j^\dagger$ for $\hat{\mathbf{x}}_j \in \hat{\mathcal{X}}$. The node function \mathbf{u} can be stored by an $N \times q$ matrix U whose i th row represents the output of \mathbf{u} at node i . We can rewrite the first term in (3) as follows:

$$\frac{1}{4} \sum_{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}} W_{ij} \|\mathbf{u}(\mathbf{x}_i) - \mathbf{u}(\mathbf{x}_j)\|_2^2 = \frac{1}{2} \langle U, LU \rangle_F \quad (5)$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius inner product for matrices. The matrix form (5) makes it easier to find the minimizer of the objective function $J_\ell(\mathbf{u}; \mathbf{y}^\dagger)$.

C. Graph-Based Active Learning

The performance of the graph learning method depends on the selection of the labeled node set $\hat{\mathcal{X}}$. The classification

accuracy varies over different selections under the same number of label nodes. A graph-based active learning approach is an iterative process developed to boost classification accuracy by strategically selecting a series of nodes on a graph $G = (\mathcal{X}, W)$ to label. Specifically, one initializes a label set \mathcal{X}_L^0 , chosen randomly in practice. At the i th iteration with the current label set $\hat{\mathcal{X}}^{i-1}$, one classifies the remaining unlabeled nodes $\mathcal{X} \setminus \hat{\mathcal{X}}^{i-1}$ via a certain graph learning scheme. Using the classification result on the set $\mathcal{X} \setminus \hat{\mathcal{X}}^{i-1}$, one calculates an acquisition function $\mathcal{A} : \mathcal{X} \setminus \hat{\mathcal{X}}^{i-1} \rightarrow \mathbb{R}$. Then, a query set $\mathcal{Q} \subset \mathcal{X} \setminus \hat{\mathcal{X}}^{i-1}$ is selected to update the current label set $\hat{\mathcal{X}}^i = \hat{\mathcal{X}}^{i-1} \cup \mathcal{Q}$.

The sequential active learning selects the query set \mathcal{Q} to be a single node with the largest acquisition function value in $\mathcal{X} \setminus \hat{\mathcal{X}}$, i.e., $\mathcal{Q} = \{\hat{\mathbf{x}}_{k^*}\}$, $\hat{\mathbf{x}}_{k^*} = \arg \max_{\mathbf{x} \in \mathcal{X} \setminus \hat{\mathcal{X}}} \mathcal{A}(\mathbf{x})$. The batch active learning selects a query set of multiple unlabeled nodes in graph G . In this article, we adopt the LocalMax [56] batch active learning approach, as it performs similar to sequential active learning but is more efficient, proportionally to the batch size. It is originally developed for the classification task of the synthetic aperture radar (SAR) images [56] and is further proved successful in the image segmentation task [50]. Specifically, in each iteration of the active learning process, the batch active learning selects the query set \mathcal{Q}_{LM} with the size $B > 1$, containing the unlabeled nodes that satisfy the local-max condition on the graph G .

This article uses graph Laplace learning [55] as the underlying graph-based classifier in the active learning process, together with two specific acquisition functions: the variance optimality (VOpt) acquisition function [51] and the model-change VOpt (MCVOpt) acquisition function [45]. Details about the formulas of these two acquisition functions are provided in the Appendix.

III. SEMISUPERVISED HSU

This section details our semisupervised HSU methods based on the linear mixing model. Specifically, given a hyperspectral data cube I of the dimension $m \times n \times p$ with p spectral channels, we reshape I into a matrix $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{p \times N}$, where $N = m \times n$ is the number of pixels. Denote a nonnegative constraint set $\Omega_{p \times q} = \{S \in \mathbb{R}^{p \times q} : S_{ij} \geq 0\}$ and a probability simplex constraint set $\Pi_{q \times N} = \{A \in \Omega_{q \times N} : \mathbf{1}_q^\top A = \mathbf{1}_N^\top\}$. We assume a linear mixing model that generates the data, i.e.,

$$X = SA + E \quad (6)$$

where $S \in \Omega_{p \times q}$ is the end-member spectrum matrix, $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N] \in \Pi_{q \times N}$ is the matrix of abundance maps of q materials, and the matrix $E \in \mathbb{R}^{p \times N}$ denotes a noise term.

In Section III-A, we describe the training data selection process by adapting the graph-based active (Section II-C) to the hyperspectral setting. Then, in Section III-B, we introduce the GLU model, which applies graph Laplace learning (Section II-B) directly to the HSU problem. Finally, in Section III-C, we propose our GRSU model that combines graph-based regularization terms with the linear mixing model (6).

A. Training Data Selection

We use the notations X and \mathcal{X} to denote a data matrix and a set of nodes, respectively. Given \mathcal{X} , we can construct a graph $G = (\mathcal{X}, W)$ according to Section II-A. Then, we apply the graph-based active learning process (Section II-C) to select a set of pixels to acquire labels for both of the proposed models, GLU (Section III-B) and GRSU (Section III-C).

Given two positive integers $m < M$, we begin the active learning process with an initial label set of m random pixels. In each iteration, we apply graph Laplace learning (Section II-B) with the current label set and calculate the acquisition function on the remaining pixels. Based on the values obtained by the acquisition function, we select a query set to be augmented to the label set and terminate this iterative process when the size of the current label set reaches M . Algorithm 1 summarizes the active learning process, and its outputs, the label dataset $\hat{\mathcal{X}} = \{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_M\}$ and the set of corresponding labels $\hat{\mathcal{A}} = \{\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_M\}$, serve as the training data for our semisupervised unmixing framework.

We want to clarify three aspects of the outputs of the active learning (Algorithm 1). First, one example of acquiring labels in Algorithm 1 is a human-in-the-loop process. Second, the ground-truth “label” $\hat{\mathbf{y}}_i \in \mathbb{R}^{q \times 1}$ for the “labeled” pixel $\hat{\mathbf{x}}_i \in \mathbb{R}^{p \times 1}$ can be either the ground-truth abundance vector or its OH pseudo-label, the latter of which can be determined by the experts for identifying the most significant end-member. According to the experimental results in Section IV, requiring the ground-truth abundance for active learning is not necessary. Third, we adopt the matrix forms of $\hat{X} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_M] \in \mathbb{R}^{p \times M}$ and $\hat{A} = [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_M] \in \mathbb{R}^{q \times M}$ of the output sets $\hat{\mathcal{X}}$ and $\hat{\mathcal{A}}$, respectively.

Given the training matrix \hat{X} , we estimate the abundance map A for the entire data matrix X . It is true that there is an overlap between \hat{X} and X , but we cannot completely trust the training labels, especially those obtained by using the OH pseudo-labels. As a result, we update the abundance map for all the pixels even though a subset of them are selected to acquire some sort of ground-truth information. Another rationale to have two separate matrices X and \hat{X} is the option to select the training pixels from one image and perform semisupervised unmixing on the other image, which falls out of the scope of this article.

B. Graph Learning Unmixing

Following the training data selection process, we obtain a training data matrix $\hat{X} \in \mathbb{R}^{p \times M}$ consisting of M labeled pixels. We concatenate \hat{X} with the original data matrix X and construct a graph \tilde{G} based on the combined data matrix $\tilde{X} = [\hat{X}, X]$ with the corresponding graph Laplacian \tilde{L} according to Section II-A. The graph Laplace learning produces the class probability, which can be regarded as the abundance map. Specifically, we adopt (5) to estimate the abundance map by projecting the graph Laplace learning solution A_{GL} onto $\Pi_{q \times N}$, i.e.,

$$A_{GL} = \arg \min_{A \in \mathbb{R}^{q \times N}} \frac{1}{2} \left\langle [\hat{A}, A]^\top, \tilde{L}[\hat{A}, A] \right\rangle_F \quad (7)$$

$$A_{GLU} = \mathcal{P}_{\Pi_{q \times N}}(A_{GL}). \quad (8)$$

Algorithm 1 Sample Labeled Pixels Through Active Learning

INPUT: dataset \mathcal{X} ; corresponding graph $G = (\mathcal{X}, W)$; initial sample number m ; total sample number M .
 OUTPUT: the label dataset $\hat{\mathcal{X}} \subset \mathcal{X}$ and the set of corresponding labels $\hat{\mathcal{A}}$.
 INITIALIZE: randomly sample m pixels as the initial label set \hat{X} ; acquire the labels for \hat{X} as \hat{A} .
 WHILE $|\hat{\mathcal{X}}| < M$ ($|\cdot|$ means cardinality) DO:
 Apply the graph Laplace learning on G based on labels of $\hat{\mathcal{X}}$ to predict labels on $\mathcal{X} \setminus \hat{\mathcal{X}}$.
 Calculate the acquisition function \mathcal{A} on $\mathcal{X} \setminus \hat{\mathcal{X}}$ based on the Laplace learning outputs.
 Select a query set \mathcal{Q} based on the acquisition function values according to the sequential active learning or LocalMax.
 Update the current label set: $\hat{\mathcal{X}} \rightarrow \hat{\mathcal{X}} \cup \mathcal{Q}$; Acquire the labels of \mathcal{Q} and update the label set $\hat{\mathcal{A}}$ accordingly.
 END WHILE

Problem (7) is the standard graph Laplace learning. Consider the block representation of L as follows:

$$\tilde{L} = \begin{bmatrix} L_{ll} & L_{lu} \\ L_{ul} & L_{uu} \end{bmatrix} \in \mathbb{R}^{(M+N) \times (M+N)} \quad (9)$$

where $L_{ll} \in \mathbb{R}^{M \times M}$ and $L_{uu} \in \mathbb{R}^{N \times N}$ are the parts of the labeled pixels \hat{X} and the unlabeled pixels X , respectively, and $L_{ul} = L_{lu}^\top \in \mathbb{R}^{N \times M}$ represents the cross interaction of graph Laplacian between \hat{X} and X . Then, the minimization problem (7) has a closed-form solution

$$A_{GL} = -\hat{A}L_{lu}L_{uu}^{-1} \quad (10)$$

which can be solved by the preconditioned conjugate gradient method because of the symmetric and semipositive definite properties of the part L_{uu} . Equation (8) is to project the output $A_{GL} \in \mathbb{R}^{q \times N}$ of the graph Laplace learning onto the set $\Pi_{q \times N}$ by a projection operator $\mathcal{P}_{\Pi_{q \times N}}$, defined by

$$\mathcal{P}_{\Pi_{q \times N}}(A) = \arg \min_{V \in \Pi_{q \times N}} \|V - A\|_F. \quad (11)$$

This projection operator can be implemented by a fast algorithm [57].

Given A_{GLU} , we can then find the optimal end-member matrix S_{GLU} that minimizes the combination of the least-squares errors of the linear mixing model (6) and the misfit of the training data, i.e.,

$$S_{GLU} = \arg \min_{S \in \Omega_{p \times q}} \frac{1}{2} \|X - SA_{GLU}\|_F^2 + \frac{\alpha^2}{2} \|\hat{X} - S\hat{A}\|_F^2 \quad (12)$$

with a weighting parameter $\alpha > 0$. Equation (12) has a closed-form solution

$$S_{GLU}^0 = \left(XA^\top + \alpha^2 \hat{X}\hat{A}^\top \right) \left(AA^\top + \alpha^2 \hat{A}\hat{A}^\top \right)^{-1} \quad (13)$$

$$S_{GLU} = \max(0, S_{GLU}^0). \quad (14)$$

Equation (14) means to take the entrywise maximum of S_{GLU}^0 and 0, i.e., replace each negative entry in S_{GLU}^0 by 0. Algorithm 2 presents the pseudo-code of our GLU method, which only involves three steps to find A_{GLU} and S_{GLU} (no iteration is needed).

C. Graph-Regularized SemiSupervised Unmixing

By assuming the linear mixing model (6), it is standard to solve the blind unmixing problem in a regularized least square

Algorithm 2 GLU

INPUT: data matrix X , training data (\hat{X}, \hat{A}) , and $\alpha > 0$.
 OUTPUT: matrices S_{GLU} and A_{GLU} .
 INITIALIZE: build a graph on $\tilde{X} = [\hat{X}, X]$ with corresponding Laplacian matrix \tilde{L} . Segment \tilde{L} into the block form (9).

1, Graph Learning Step:

$$A_{GL} = -\hat{A}L_{lu}L_{uu}^{-1}.$$

2, Projection:

$$A_{GLU} = \mathcal{P}_{\Pi_{q \times N}}(A_{GL}).$$

3, Estimate the endmember spectrum matrix:

$$S_{GLU}^0 = \left(XA^\top + \alpha^2 \hat{X}\hat{A}^\top \right) \left(AA^\top + \alpha^2 \hat{A}\hat{A}^\top \right)^{-1},$$

$$S_{GLU} = \max(0, S_{GLU}^0).$$

form

$$\arg \min_{\substack{S \in \Omega_{p \times q}, \\ A \in \Pi_{q \times N}}} \frac{1}{2} \|X - SA\|_F^2 + \lambda \mathcal{J}(A) \quad (15)$$

with a positive weighting parameter λ . The term $(1/2)\|X - SA\|_F^2$ is a least-squares misfit between the matrix product SA and the data measurement X , while $\mathcal{J}(A)$ is a regularization term of the abundance matrix A . We impose the nonnegative constraints of S and A as well as a sum-to-one constraint of A .

Following the graph Dirichlet energy [58], [59], we consider a graph Laplacian regularization, formulated by:

$$\mathcal{J}_1(A) = \frac{1}{4} \sum_{i,j=1}^N \|\mathbf{a}_i - \mathbf{a}_j\|_2^2 w(\mathbf{x}_i, \mathbf{x}_j) \quad (16)$$

where the weight function $w : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ is defined by

$$w(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{\sigma_i \sigma_j}\right)$$

with the angular distance $d(\mathbf{x}_i, \mathbf{x}_j)$ between node \mathbf{x}_i and \mathbf{x}_j , and σ_i and σ_j defined in (1) and (2), respectively.

In addition, we further develop a semisupervised graph regularization term $\mathcal{J}_2(A, \hat{A}, \hat{X})$ that includes the label information \hat{X} and \hat{A} , that is,

$$\mathcal{J}_2(A; \hat{A}, \hat{X}) = \frac{1}{4} \sum_{i=1}^N \sum_{j=1}^M \|\mathbf{a}_i - \hat{\mathbf{y}}_j\|_2^2 w(\mathbf{x}_i, \hat{\mathbf{x}}_j). \quad (17)$$

We consider the sum of both terms $\mathcal{J}_1(A)$ and $\mathcal{J}_2(A; \hat{A}, \hat{X})$ as the regularization $\mathcal{J}(A)$. As \mathcal{J}_1 and \mathcal{J}_2 have the same form that

leads to the same scale, we assign the equal weight of them when formulating \mathcal{J} , i.e., $\mathcal{J}(A) = \mathcal{J}_1(A) + \mathcal{J}_2(A; \hat{A}, \hat{X})$.

Putting (15)–(17) together with the label information (\hat{X}, \hat{A}) obtained by the active learning approach, we propose a GRSU model to simultaneously estimate the abundance map A_{GRSU} and the end-member matrix S_{GRSU} , that is,

$$S_{\text{GRSU}}, A_{\text{GRSU}} = \arg \min_{\substack{S \in \Omega_{p \times q}, \\ A \in \Pi_{q \times N}}} \frac{1}{2} \|X - SA\|_F^2 + \frac{\alpha^2}{2} \|\hat{X} - S\hat{A}\|_F^2 + \lambda \mathcal{J}_1(A) + \lambda \mathcal{J}_2(A; \hat{A}, \hat{X}) \quad (18)$$

with two positive parameters α and λ . We define the indicator function

$$\chi_{\Delta}(Z) = \begin{cases} 0, & Z \in \Delta \\ \infty, & \text{otherwise} \end{cases} \quad (19)$$

to rewrite the minimization problem (18) into an unconstrained formulation as follows:

$$\min_{A, S} \frac{1}{2} \|X - SA\|_F^2 + \frac{\alpha^2}{2} \|\hat{X} - S\hat{A}\|_F^2 + \lambda \mathcal{J}_1(A) + \lambda \mathcal{J}_2(A; \hat{A}, \hat{X}) + \chi_{\Omega_{p \times q}}(S) + \chi_{\Pi_{q \times N}}(A). \quad (20)$$

We apply the alternating direction method of multipliers (ADMM) [60] to solve the unconstrained problem (20). In particular, we introduce two auxiliary variables $T \in \mathbb{R}^{p \times q}$ and $B \in \mathbb{R}^{q \times N}$ to express problem (20) equivalently as follows:

$$\begin{aligned} \min_{A, B, S, T} & \frac{1}{2} \|X - TA\|_F^2 + \frac{\alpha^2}{2} \|\hat{X} - T\hat{A}\|_F^2 + \lambda \mathcal{J}_1(B) \\ & + \lambda \mathcal{J}_2(B; \hat{A}, \hat{X}) + \chi_{\Omega_{p \times q}}(S) + \chi_{\Pi_{q \times N}}(A) \\ \text{s.t. } & A = B, S = T. \end{aligned} \quad (21)$$

The augmented Lagrangian of (21) is written as follows:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \|X - TA\|_F^2 + \frac{\alpha^2}{2} \|\hat{X} - T\hat{A}\|_F^2 + \lambda \mathcal{J}_1(B) \\ & + \lambda \mathcal{J}_2(B; \hat{A}, \hat{X}) + \chi_{\Omega_{p \times q}}(S) + \chi_{\Pi_{q \times N}}(A) \\ & + \frac{\rho}{2} \|A - B + \bar{B}\|_F^2 + \frac{\gamma}{2} \|S - T + \bar{T}\|_F^2 \end{aligned}$$

with dual variables \bar{B} and \bar{T} and two positive constants ρ and γ . One benefit of ADMM is that it turns the joint minimization problem (21) into four subproblems that are associated A , B , S , and T separately. In each iteration, we iterate as follows:

$$\begin{aligned} T & \leftarrow \arg \min_{T \in \mathbb{R}^{p \times q}} \frac{1}{2} \|[X, \alpha \hat{X}] - T[A, \alpha \hat{A}]\|_F^2 + \frac{\gamma}{2} \|S - T + \bar{T}\|_F^2 \\ S & \leftarrow \arg \min_{S \in \Omega_{p \times q}} \frac{\gamma}{2} \|S - T + \bar{T}\|_F^2 \\ A & \leftarrow \arg \min_{A \in \Pi_{q \times N}} \frac{1}{2} \|X - TA\|_F^2 + \frac{\rho}{2} \|A - B + \bar{B}\|_F^2 \\ B & \leftarrow \arg \min_{B \in \mathbb{R}^{q \times N}} \lambda \mathcal{J}_1(B) + \lambda \mathcal{J}_2(B; \hat{A}, \hat{X}) + \frac{\rho}{2} \|A - B + \bar{B}\|_F^2 \\ \bar{B} & \leftarrow \bar{B} - A + B \\ \bar{T} & \leftarrow \bar{T} - S + T. \end{aligned} \quad (22)$$

The A , S , and T subproblems are the same as in the blind unmixing paper [29]; thus, the details are omitted.

As for the B subproblem, we write it explicitly by using $B = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N]$

$$\begin{aligned} B & = \arg \min_{B \in \mathbb{R}^{q \times N}} \mathcal{J}_1(B) + \mathcal{J}_2(B; \hat{A}, \hat{X}) + \frac{\rho}{2\lambda} \|A - B + \bar{B}\|_F^2 \\ & = \arg \min_{B \in \mathbb{R}^{q \times N}} \frac{1}{4} \sum_{i,j=1}^N \|\mathbf{b}_i - \mathbf{b}_j\|_2^2 w(\mathbf{x}_i, \mathbf{x}_j) \\ & \quad + \frac{1}{4} \sum_{i=1}^N \sum_{j=1}^M \|\mathbf{b}_i - \hat{\mathbf{y}}_j\|_2^2 w(\mathbf{x}_i, \hat{\mathbf{x}}_j) \\ & \quad + \frac{1}{4} \sum_{i,j=1}^M \|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j\|_2^2 w(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j) + \frac{\rho}{2\lambda} \|A - B + \bar{B}\|_F^2. \end{aligned} \quad (23)$$

Note that we add the term $(1/4) \sum_{i,j=1}^M \|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j\|_2^2 w(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j)$ in (23) that does not affect the minimization over B , but rather turns the B subproblem (23) into a regularized graph Laplacian learning problem (see Section II-B). Specifically, we consider the graph \tilde{G} built on the combined data matrix $\tilde{X} = [\hat{X}, X]$ with the graph Laplacian matrix \tilde{L} . Then, (23) is equivalent to

$$B = \arg \min_{B \in \mathbb{R}^{q \times N}} \frac{1}{2} \left\langle [\hat{A}, B]^\top, \tilde{L}[\hat{A}, B]^\top \right\rangle_F + \frac{\rho}{2\lambda} \|A - B + \bar{B}\|_F^2. \quad (24)$$

Using the block representation of \tilde{L} in (9), we have a closed-form solution to (24) as follows:

$$B = \left(-\hat{A}L_{lu} + \frac{\rho}{\lambda}(A + \bar{B}) \right) \left(L_{uu} + \frac{\rho}{\lambda}I_N \right)^{-1} \quad (25)$$

where I_N is an $N \times N$ identity matrix.

The semisupervised unmixing (GRSU) method is summarized in Algorithm 3. Its initial values of A^0 and S^0 are obtained by GLU (Section III-B) that outputs A_{GLU} and S_{GLU} .

IV. EXPERIMENTS AND RESULTS

In this section, we conduct extensive experiments to demonstrate the performance of the proposed unmixing models. All codes of our proposed methods and following experiments are available on our GitHub repository.¹ Specifically, in Section IV-A, we compare our semisupervised methods (GLU and GRSU) with the state-of-the-art (unsupervised) blind unmixing methods, followed by a discussion of our semisupervised unmixing methods with respect to different numbers of training pixels in Section IV-B. In Section IV-C, we test the robustness of various methods by adding different amounts of Gaussian white noise to the HSI. We test on four standard hyperspectral image datasets, described as follows.

- 1) *Jasper Ridge*: The Jasper Ridge dataset [61] is a hyperspectral image of the size 100×100 with 198 channels. Originally, it had 224 hyperspectral channels spanning from 380 to 2500 nm, and 26 channels are removed as a preprocessing step due to dense water vapor and atmospheric effects. Four end-members are latent: Tree, Water, Dirt, and Road.

¹<https://github.com/wispcarey/Nearly-Blind-Hyperspectral-Unmixing>

Algorithm 3 Semisupervised Unmixing

INPUT: data matrix X , label information (\hat{X}, \hat{A}) , parameters $\alpha, \lambda, \gamma, \rho$, maximum iteration I_{\max} , and error tolerance ϵ .
 CONSTRUCT: a graph \tilde{G} on $\tilde{X} = [\hat{X}, X]$ (by Section II-A) with the graph Laplacian \tilde{L} and its block form of (9)
 INITIALIZE: $S^0 = S_{\text{GLU}}$ and $A^0 = A_{\text{GLU}}$ (by GLU Algorithm 2); $B^0 = A^0$, $\bar{B}^0 = 0$, $\bar{T} = 0$, $\text{Err} = 1$, and $i = 0$.
 WHILE $i < I_{\max}$ AND $\text{Err} > \epsilon$, DO:
 $T^{i+1} = \left(X(A^i)^\top + \alpha^2 \hat{X}(\hat{A})^\top + \gamma(S^i + \bar{T}^i) \right) \left(A^i(A^i)^\top + \alpha^2 \hat{A}(\hat{A})^\top + \gamma I_q \right)^{-1}$.
 $S^{i+1} = \max(T^{i+1} - \bar{T}^i, 0)$.
 $A^{i+1} = \mathcal{P}_{\Pi_{q \times N}} \left(((S^{i+1})^\top S^{i+1} + \rho I_q)^{-1} ((S^{i+1})^\top X + \rho(B^i - \bar{B}^i)) \right)$.
 $B^{i+1} = \left(-\hat{A} L_{ul} + \frac{\rho}{\lambda} (A^{i+1} + \bar{B}^i) \right) (L_{uu} + \frac{\rho}{\lambda} I_N)^{-1}$
 $\bar{B}^{i+1} = \bar{B}^i + (A^{i+1} - B^{i+1})$.
 $\bar{T}^{i+1} = \bar{T}^i + (S^{i+1} - T^{i+1})$.
 $i \leftarrow i + 1$ and $\text{Err} = \max((\|S^{i+1} - S^i\|_F) / (\|S^i\|_F), (\|A^{i+1} - A^i\|_F) / (\|A^i\|_F))$
 END WHILE
 OUTPUT: $S_{\text{GRSU}} = S^i$, $A_{\text{GRSU}} = A^i$

- 2) *Samson*: The Samson dataset [61] is of the size $95 \times 95 \times$ with 156 channels that span from 401 to 889 nm. Three end-members are latent: Soil, Tree, and Water. Note that a different ground truth is considered in [35], while we use the original ground-truth information.
- 3) *Urban4*: The Urban dataset [61] is of the size 307×307 with 162 channels. Each pixel of this image corresponds to a $2\text{-m} \times 2\text{-m}$ area. The original 221 channels span from 400 to 2500 nm. There are three versions of the ground truth, which contain four, five, and six end-members. Here, we use the version of four end-members, labeled as Asphalt, Grass, Tree, and Roof.
- 4) *Apex*: The Apex dataset² [62] is a hyperspectral image of the size 111×122 with 285 bands spanning from 413 to 2420 nm. Four end-members are latent in this data: Road, Tree, Roof, and Water.

We apply two metrics, root-mean-square error (RMSE) and spectral angle distance (SAD), to evaluate the quality of the abundance matrix A and the end-member spectrum matrix S , respectively. RMSE and SAD are defined as follows:

$$\text{RMSE}(A, A^{\text{gt}}) = 100 \times \sqrt{\frac{1}{pn} \|A - A^{\text{gt}}\|_F^2} \quad (26)$$

$$\text{SAD}(S, S^{\text{gt}}) = \frac{180}{\pi} \times \frac{1}{p} \sum_{i=1}^p \arccos \left(\frac{\langle \mathbf{s}_i, \mathbf{s}_i^{\text{gt}} \rangle}{\|\mathbf{s}_i\|_2 \|\mathbf{s}_i^{\text{gt}}\|_2} \right) \quad (27)$$

where A and S are the fitted matrices, A^{gt} and S^{gt} are the ground truth, and \mathbf{s}_i denotes the i th column of the matrix S .

A. Method Comparison

We compare our semisupervised methods with five state-of-the-art unsupervised unmixing methods, namely, FCLSU [10], GLNMF [26], QMV [20], graph total variation Merriman–Bence–Osher scheme (GTVMBOS) [30], minimum simplex convolutional network (MSC) [35], and endmember-guided unmixing network (EGU) [36]. The first three methods

(GLNMF, QMV, and GTVMBO) initialize with the output of FCLSU [10]. MSC and EGU are neural network methods. Note that the EGU method supports the use of either the ground-truth end-member spectrum matrix S or the estimated matrix S using vertex component analysis (VCA) [63]. We use VCA to estimate S , which serves as an input for the pixelwise EGU-net method. In the following experiments, GLNMF [26], QMV [20], and GTVMBO [30] are executed in MATLAB and conducted on an Intel i9-9900K CPU, while MSC [35] and EGU [36] are implemented in Python and conducted on a NVIDIA 2080 Ti GPU.

Table I shows the computation times of various methods applied to each dataset. Our semisupervised method GLU is much faster than the neural network methods approaches (MSC and EGU) and is comparable to traditional unsupervised methods (GLNMF, QMV, and GTVMBO) in terms of computation time. As the GRSU method requires solving the graph Laplace learning problem in each iteration, it does require more computation times compared with the regularization-based methods, while it is still faster than neural network methods.

For labels used in our semisupervised framework, we consider the exact (EXT) abundance map and the OH pseudo-label. The latter (OH) can be obtained by thresholding the EXT abundance map or an expert identifying the most significant end-member, which is more practical than the former (EXT) in real circumstances.

For each experiment, the active learning process starts with only one random pixel per material as an initial label set and terminates until around 0.4% of the total pixels are sampled according to the active learning algorithm (Algorithm 1). We fix $\epsilon = 10^{-3}$, $I_{\max} = 1000$, and $K = 50$ in the KNN graph construction (Section II-A). Note that K is chosen to be relatively small for computational benefits, while ensuring the connectivity of the constructed graph G , which is required in the calculation of the acquisition function (28). We select the optimal combination of the parameters α, λ, γ , and ρ in the range of $\alpha \in \{10, 20, 50, 100\}$, $\lambda \in \{10^i, 5 \times 10^i | i = 0, 1, 2, 3\}$, and $\gamma = \rho \in \{10^i | i = -2, -1, 0, 1, 2\}$ that yields the smallest sum of RMSE and SAD using 10% of randomly selected pixels as

²<https://github.com/BehnoodRasti/MiSiCNet>

TABLE I
COMPUTATION TIMES FOR DIFFERENT METHODS

Dataset	Unsupervised Methods					Our Semi-supervised Methods			
	GLNMF [26]	QMV [20]	GTVMBO [30]	MSC [35]	EGU [36]	GLU-OH	GLU-EXT	GRSU-OH	GRSU-EXT
Jasper	8.81s*	2.51s*	2.77s*	112.09s	51.36s	2.38s	2.45s	3.07s	2.92s
Samson	3.89s*	1.40s*	0.51s*	95.32s	43.58s	1.52s	1.47s	10.42s	12.92s
Urban4	67.31s*	23.57s*	14.60s*	974.05s	583.16s	22.29s	22.37s	201.89s	294.71s
Apex	26.47s*	2.12s*	0.59s*	166.03s	87.84s	3.81s	3.97s	20.46s	24.04s

This table presents the computation times of various methods applied to each dataset. Specifically, the methods of GLNMF [26], QMV [20], and GTVMBO [30] are executed in MATLAB, with their times being indicated by an asterisk (*) while the other methods are implemented in Python. The computation time is measured in seconds. In each row, the best computation times from the unsupervised and semi-supervised methods are highlighted in bold. Regarding computational time, the proposed methods (GLU and GRSU) run much faster than the neural network methods approaches (MSC and EGU), and are comparable to traditional unsupervised methods (GLNMF, QMV, and GTVMBO).

TABLE II
TRAINING INFORMATION FOR EACH DATASET

Dataset	Parameters				Training Data Info			
	α	λ	γ	ρ	Acq Fun	Training pixels	Training Percentage	Num Each Class
Jasper	10	1	1	1	MCVOPT	44	0.44%	13, 7, 16, 8
Samson	20	50	0.1	0.1	VOPT	36	0.40%	13, 14, 9
Urban4	50	500	0.1	0.1	VOPT	364	0.38%	138, 105, 72, 49
Apex	10	50	1	1	VOPT	54	0.40%	7, 24, 13, 10

Parameter choices and training data information for our GLU and GRSU models. Parameters α , λ , γ , ρ are all associated with GRSU, while GLU only involves one parameter α (same as the one used in GRSU). “Acq Fun” means the acquisition function applied in the active learning process. “Training Pixels” means the number of labeled pixels used for the training process. “Training Percentage” means the percentage of labeled pixels to all pixels. “Num Each Class” means the number of labeled pixels of each endmember.

TABLE III
RMSE BETWEEN THE ESTIMATED ABUNDANCE MATRIX AND THE GROUND TRUTH

Dataset	Class	Unsupervised Methods					Our Semi-supervised Methods (0.4% training)			
		GLNMF [26]	QMV [20]	GTVMBO [30]	MSC [35]	EGU [36]	GLU-OH	GLU-EXT	GRSU-OH	GRSU-EXT
Jasper	Tree	10.04	10.59	18.70	12.34	11.25	7.09	10.91	4.04	6.48
	Water	14.06	4.18	7.38	7.30	6.22	7.33	11.40	4.39	4.27
	Dirt	14.32	10.93	17.10	15.35	14.59	8.64	13.04	6.48	6.85
	Road	51.82	14.74	15.64	52.09	32.84	14.78	15.75	6.99	6.34
	Overall	18.79	9.51	15.16	18.24	13.78	8.37	12.03	5.10	5.93
Samson	Soil	22.51	20.28	8.01	20.05	30.73	4.37	5.73	5.18	4.10
	Tree	31.29	24.92	16.10	28.90	29.56	7.51	6.02	6.35	4.38
	Water	16.26	16.63	9.19	16.45	15.58	11.06	4.70	11.32	4.89
	Overall	25.21	21.48	12.19	23.32	27.08	7.81	5.61	7.66	4.43
Urban4	Asphalt	33.89	21.58	19.19	18.19	21.4	14.48	7.99	14.55	8.32
	Grass	10.15	29.07	10.44	31.61	45.83	7.61	6.86	7.47	6.93
	Tree	15.82	25.65	13.87	31.37	14.32	7.78	9.51	7.88	9.67
	Roof	17.53	31.18	23.49	38.00	18.86	9.93	10.65	9.78	10.59
	Overall	22.13	26.17	15.71	28.58	30.77	10.48	8.30	10.49	8.46
Apex	Road	23.13	25.34	57.78	22.77	43.61	11.35	17.40	10.16	13.68
	Tree	13.51	19.99	32.29	14.08	25.27	11.33	12.22	11.07	9.44
	Roof	29.15	25.65	46.64	19.91	33.67	16.67	12.69	17.31	13.71
	Water	16.81	8.70	21.63	5.50	11.56	14.85	14.43	15.08	14.85
	Overall	19.28	20.09	37.10	15.27	27.49	13.34	13.44	13.35	12.15

Comparison results in terms of $RMSE(A, A^{gt})$ for the abundance maps: four of our semi-supervised methods (with around 0.4% of labeled pixels) are compared with five (unsupervised) blind unmixing methods on four publicly available datasets. For each row, the best results of unsupervised methods and our semi-supervised methods are bolded, respectively. The best of our methods achieves nearly 50% improvements over the unsupervised ones in most cases.

a validation set. We list the parameter choices, the amount of training data, and the acquisition function (Acq Fun) in the active learning approach for each dataset in Table II.

Tables III and IV report the results of $RMSE(A, A^{gt})$ values and $SAD(S, S^{gt})$ values, respectively. For our methods, the “-OH” suffix refers to training on the one-hot pseudo-labels, while the “-EXT” suffix refers to training on the exact abundance maps. In both tables, we highlight the best results in our four semisupervised methods and the other five unsupervised methods, separately. By comparing the bold results in each row, the best of our methods achieves nearly 50% improvements over the competing unsupervised methods

in most cases. There are only a few exceptions. For example, EGU has an outstanding performance of the SAD on the Apex dataset, which is slightly better than ours.

In addition, we observe that the winner of the five supervised methods scatters over Tables III and IV. For example, QMV and GTVMBO perform the best on Jasper Ridge and Urban4, respectively, for both abundance and end-member estimations. For the Samson dataset, GTVMBO attains the best RMSE, while EGU has the best SAD. For Apex, MSC has the best RMSE, while EGU achieves the best SAD. Our methods, on the other hand, yield consistent performance in that GRSU is generally better than GLU. One exception is the

TABLE IV
SAD BETWEEN THE ESTIMATED SPECTRUM MATRIX AND THE GROUND TRUTH

Dataset	Class	Unsupervised Methods					Our Semi-supervised Methods (0.4% training)			
		GLNMF [26]	QMV [20]	GTVMBO [30]	MSC [35]	EGU [36]	GLU-OH	GLU-EXT	GRSU-OH	GRSU-EXT
Jasper	Tree	5.63	2.66	13.95	2.48	8.49	7.49	2.50	2.00	4.70
	Water	3.60	5.02	22.43	16.60	14.64	22.21	22.66	5.04	22.27
	Dirt	6.53	2.50	8.37	3.80	6.68	3.99	1.80	1.89	2.54
	Road	43.94	3.96	6.61	18.88	5.16	2.41	2.05	1.28	2.20
	Overall	14.92	3.55	12.83	10.44	8.74	9.03	7.25	2.55	7.92
Samson	Soil	1.37	2.48	2.59	23.13	1.35	2.07	1.87	1.44	2.32
	Tree	1.62	3.04	5.53	2.03	2.29	4.12	2.54	3.20	2.56
	Water	10.48	32.94	21.40	45.95	8.62	9.50	30.96	2.41	31.45
	Overall	4.50	12.82	9.83	23.71	4.09	5.24	11.79	2.36	12.11
Urban4	Asphalt	7.36	150.64	6.98	43.06	7.62	6.35	3.02	6.25	3.08
	Grass	10.85	23.49	6.66	22.39	37.45	3.49	3.82	3.35	3.74
	Tree	9.11	7.52	1.58	5.67	4.86	7.02	3.80	6.35	4.39
	Roof	44.15	3.85	44.23	2.59	45.89	3.26	3.89	3.22	4.16
	Overall	17.87	46.37	14.87	18.43	23.96	5.04	3.64	4.79	3.83
Apex	Road	29.38	8.39	40.88	14.90	5.61	7.68	14.95	4.25	17.61
	Tree	6.55	11.86	24.21	7.88	7.41	10.15	5.08	8.03	6.62
	Roof	6.23	6.26	5.65	10.10	4.86	9.59	6.47	6.82	4.30
	Water	9.70	106.13	51.46	43.85	12.73	30.73	11.33	12.62	17.11
	Overall	12.96	33.16	30.55	19.19	7.65	14.54	9.46	7.94	11.41

Comparison results in terms of SAD(S, S^{gt}) for the endmember spectrum matrices S : four of our semi-supervised methods (with around 0.4% of labeled pixels) are compared with five (unsupervised) blind unmixing methods on four publicly available datasets. For each row, the best results of unsupervised methods and our semi-supervised methods are bolded, respectively.

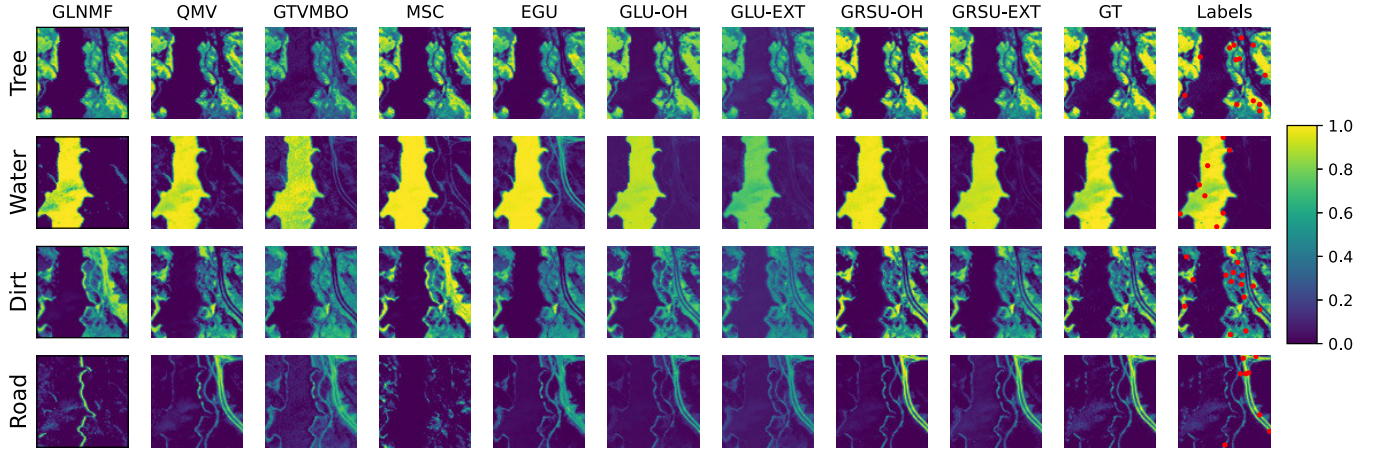


Fig. 2. Abundance maps estimated by different methods on the Jasper Ridge dataset. Each row of the plot matrix corresponds to an end-member of the dataset, including Tree, Water, Dirt, and Road. The first five columns are (unsupervised) blind unmixing methods, and the following four columns are our semisupervised methods. The last two columns are the ground truth and the label pixels selected by active learning (red dots) for our GLU and GRSU methods. Each red dot corresponds to a labeled pixel, which is enlarged for visual illustration. Notice that the Road abundance map contains fine structures that are easily smeared out by other methods, while active learning can successfully identify pixels with high abundance values for Road to acquire labels.

Urban4 dataset, in which GRSU-EXT is slightly worse than GLU-EXT. Since GLU is the initialization of GRSU, we can see improvements of GRSU after the ADMM iterations, which are particularly significant on the SAD values of the Jasper Ridge, Samson, and Apex datasets. Furthermore, training on OH pseudo-labels sometimes has a better performance than training on the EXT abundance map, which implies that it is not necessary to require the EXT abundance maps for the training process of our approaches.

We arrange the estimated abundance maps and end-members in pairs for Jasper Ridge, Samson, Urban4, and Apex datasets, sequentially, showing in Figs. 2–9. Note that we normalize each end-member (column) in the spectrum matrix S to have the unit norm, i.e., $\|s_i\|_2 = \|s_i^{\text{gt}}\|_2 = 1$. The labeled pixels that are selected by active learning are indicated in red dots on the ground-truth abundance map.

The active learning approach can identify the distribution of each end-member by selecting a few representatives. Take the Road in the Jasper Ridge dataset for an example. As illustrated in the last row of Fig. 2, the abundance map for the Road contains fine structures that are easily smeared out by other methods, while active learning can successfully identify pixels with high abundance values of this end-member to acquire labels. With those sampled labeled pixels, the estimated abundance maps' quality increases significantly compared with unsupervised methods. Another evident example is the end-member Roof in the Urban dataset (the last row of Fig. 6). Both GLU-OH and GRSU-OH well preserve the contrast of the rectangular rooftop in the Roof abundance.

In conclusion, our methods demonstrate a significant improvement of approximately 50% in unmixing performance as measured by the RMSE of the abundance maps and

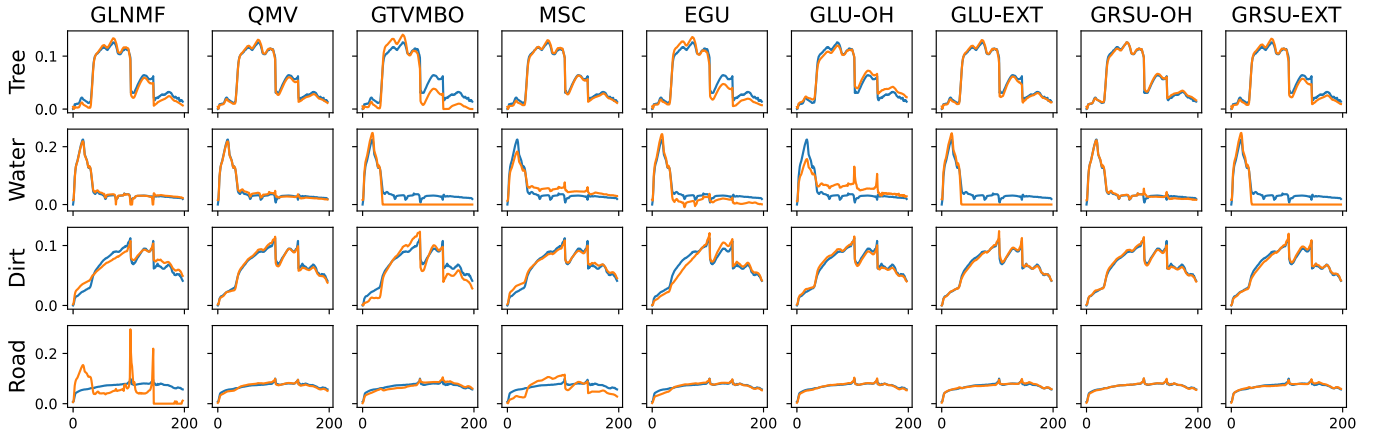


Fig. 3. End-member matrix estimated by different methods (in orange) on the Jasper Ridge dataset with the ground truth (in blue). All the end-member vectors are normalized to have the unit norm. Each row of the plot matrix corresponds to an end-member of the dataset, including Tree, Water, Dirt, and Road. The first five columns are (unsupervised) blind unmixing methods, and the following four columns are our semisupervised methods.

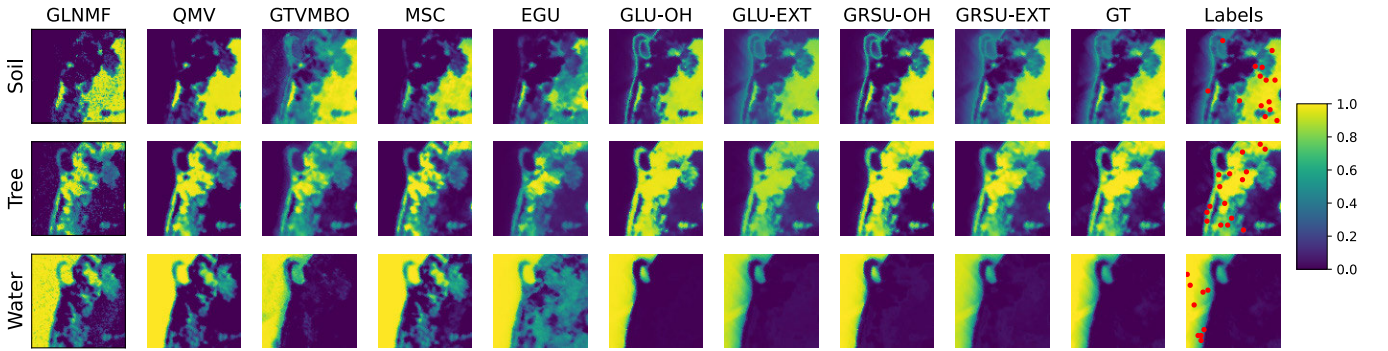


Fig. 4. Abundance maps estimated by different methods on the Samson dataset. Each row of the plot matrix corresponds to an end-member of the dataset, including Soil, Tree, and Water. The first five columns are (unsupervised) blind unmixing methods, and the following four columns are our semisupervised methods. The last two columns are the ground truth and the label pixels selected by active learning (red dots) for our GLU and GRSU methods. Each red dot corresponds to a labeled pixel, which is enlarged for visual illustration.

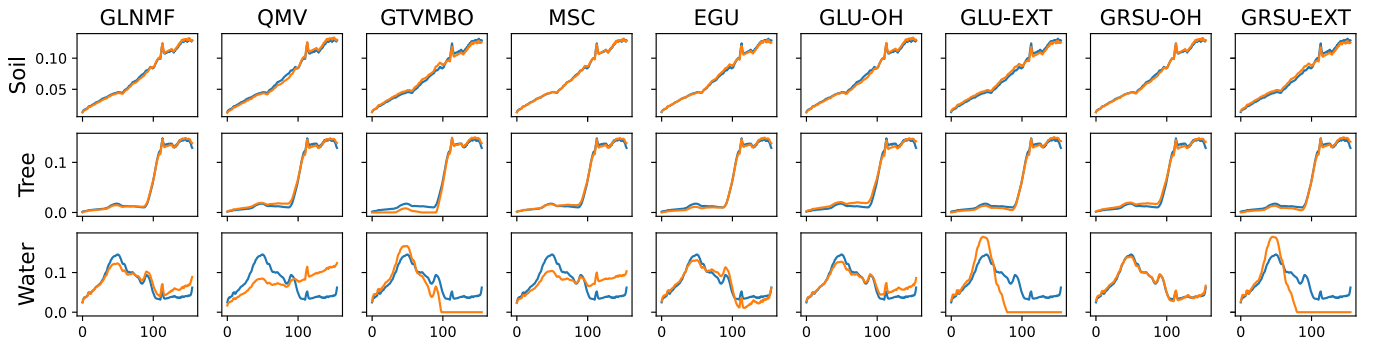


Fig. 5. End-member matrix estimated by different methods (in orange) on the Samson dataset with the ground truth (in blue). All the end-member vectors are normalized to have the unit norm. Each row of the plot matrix corresponds to an end-member of the dataset, including Soil, Tree, and Water. The first five columns are (unsupervised) blind unmixing methods, and the following four columns are our semisupervised methods.

the SAD of the mixing matrices, requiring only a minimum amount of supervision (e.g., 0.4% labeled pixels). It is important to note that EXT abundance maps are not needed during the training process; instead, we leverage the practical and readily obtainable OH labels. Furthermore, while our GRSU methods may exhibit slower computation times due to the iterative nature of our optimization approach, they still maintain a speed advantage over neural network methods, such

as MSC and EGU. This not only underlines the efficiency of our methods but also represents a competitive balance between computational speed and unmixing performance.

B. Discussion on the Number of Training Pixels

We discuss the influence of the number of labels sampled by active learning on our semisupervised methods, GLU-OH,

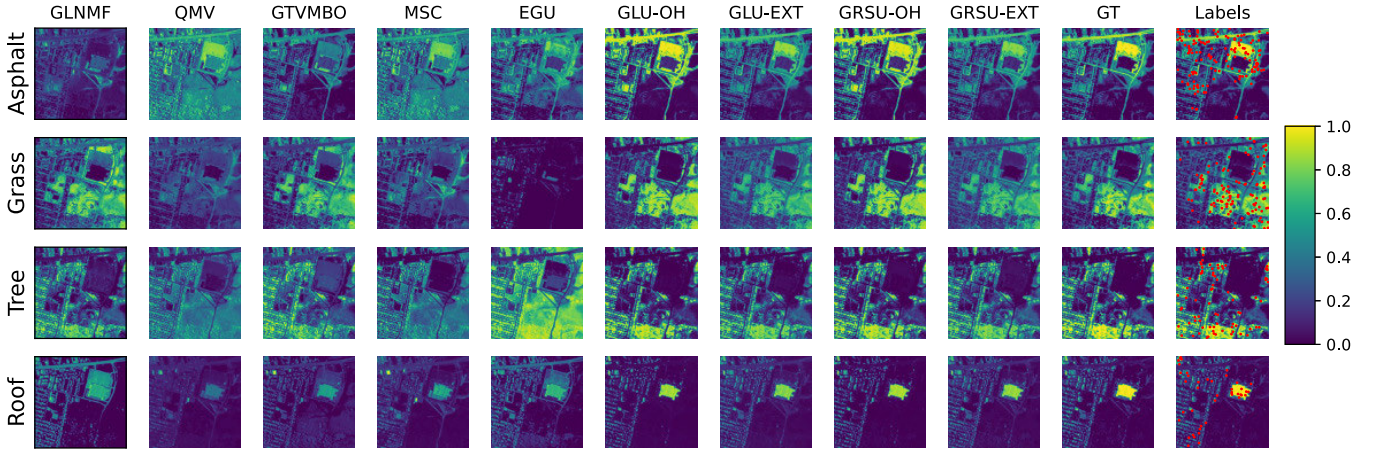


Fig. 6. Abundance maps estimated by different methods on the Urban4 dataset. Each row of the plot matrix corresponds to an end-member of the dataset, including Asphalt, Grass, Tree, and Roof. The first five columns are (unsupervised) blind unmixing methods, and the following four columns are our semisupervised methods. The last two columns are the ground truth and the label pixels selected by active learning (red dots) for our GLU and GRSU methods. Each red dot corresponds to a labeled pixel, which is enlarged for visual illustration. Note that both GLU-OH and GRSU-OH well preserve the contrast of the rectangular rooftop in the Roof abundance.

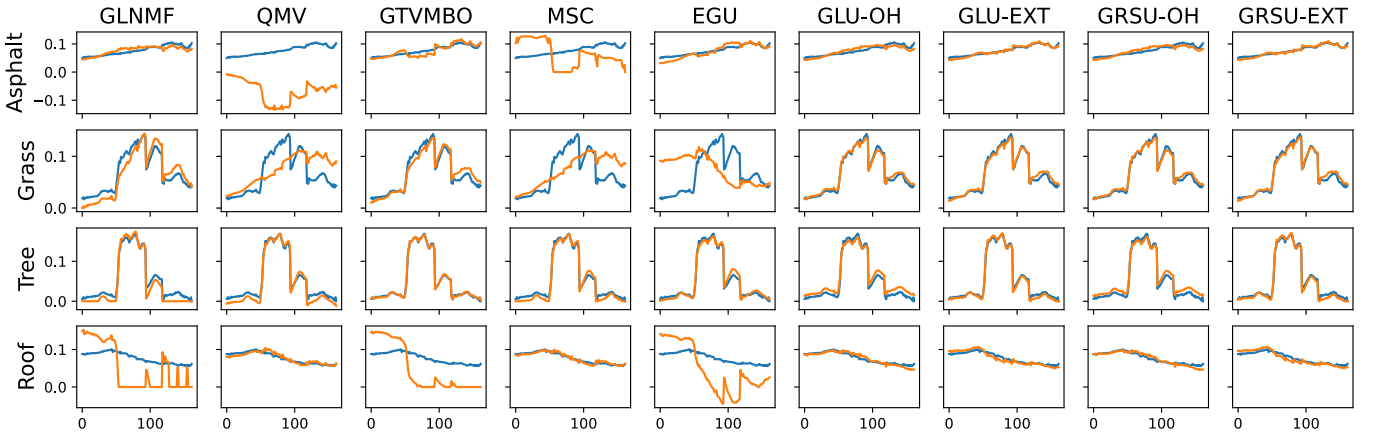


Fig. 7. End-member matrix estimated by different methods (in orange) on the Urban4 dataset with the ground truth (in blue). All the end-member vectors are normalized to have the unit norm. Each row of the plot matrix corresponds to an end-member of the dataset, including Asphalt, Grass, Tree, and Roof. The first five columns are (unsupervised) blind unmixing methods, and the following four columns are our semisupervised methods.

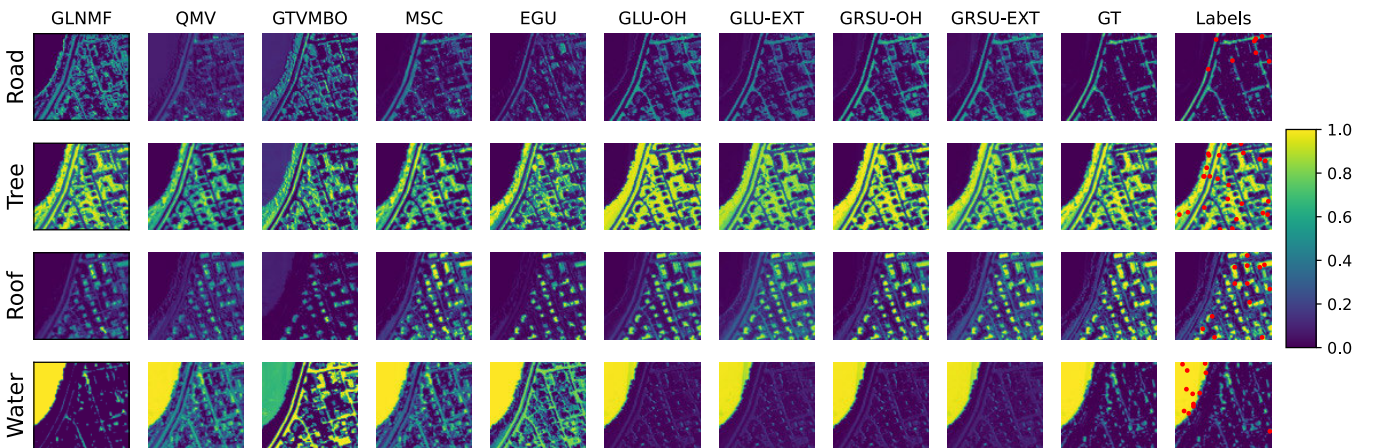


Fig. 8. Abundance maps estimated by different methods on the Apex dataset. Each row of the plot matrix corresponds to an end-member of the dataset, including Road, Tree, Roof, and Water. The first five columns are (unsupervised) blind unmixing methods, and the following four columns are our semisupervised methods. The last two columns are the ground truth and the label pixels selected by active learning (red dots) for our GLU and GRSU methods. Each red dot corresponds to a labeled pixel, which is enlarged for visual illustration.

GLU-EXT, GRSU-OH, and GRSU-EXT. The active learning process starts with one random pixel per material as an initial label set, followed by Algorithm 1 (the active learning

algorithm), until 5% of the total pixels are reached. We conduct experiments only on the Jasper Ridge and the Apex datasets for demonstration purposes. The corresponding parameters

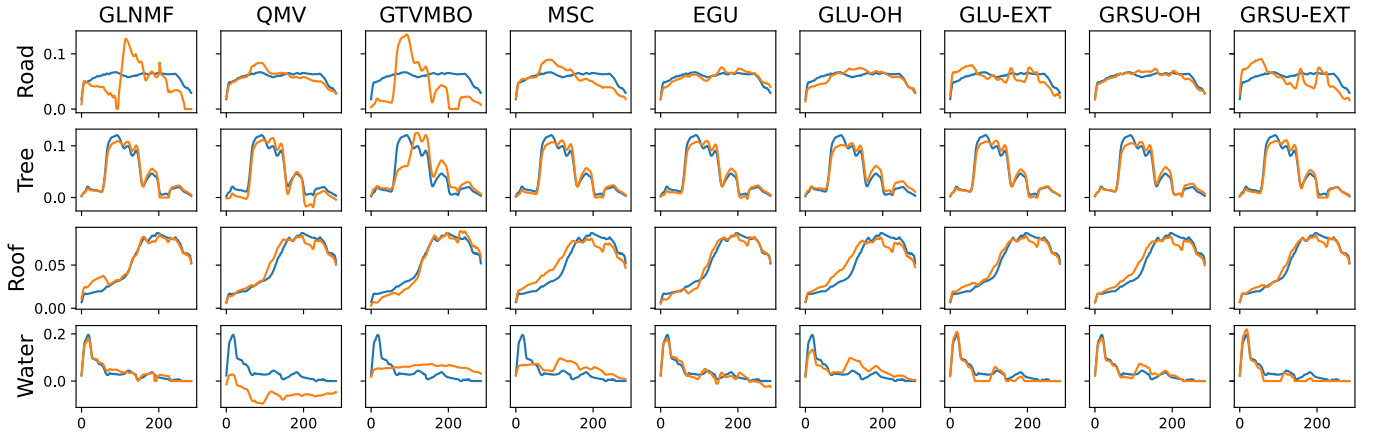


Fig. 9. End-member matrix estimated by different methods (in orange) on the Apex dataset with the ground truth (in blue). All the end-member vectors are normalized to have the unit norm. Each row of the plot matrix corresponds to an end-member of the dataset, including Road, Tree, Roof, and Water. The first five columns are (unsupervised) blind unmixing methods, and the following four columns are our semisupervised methods.

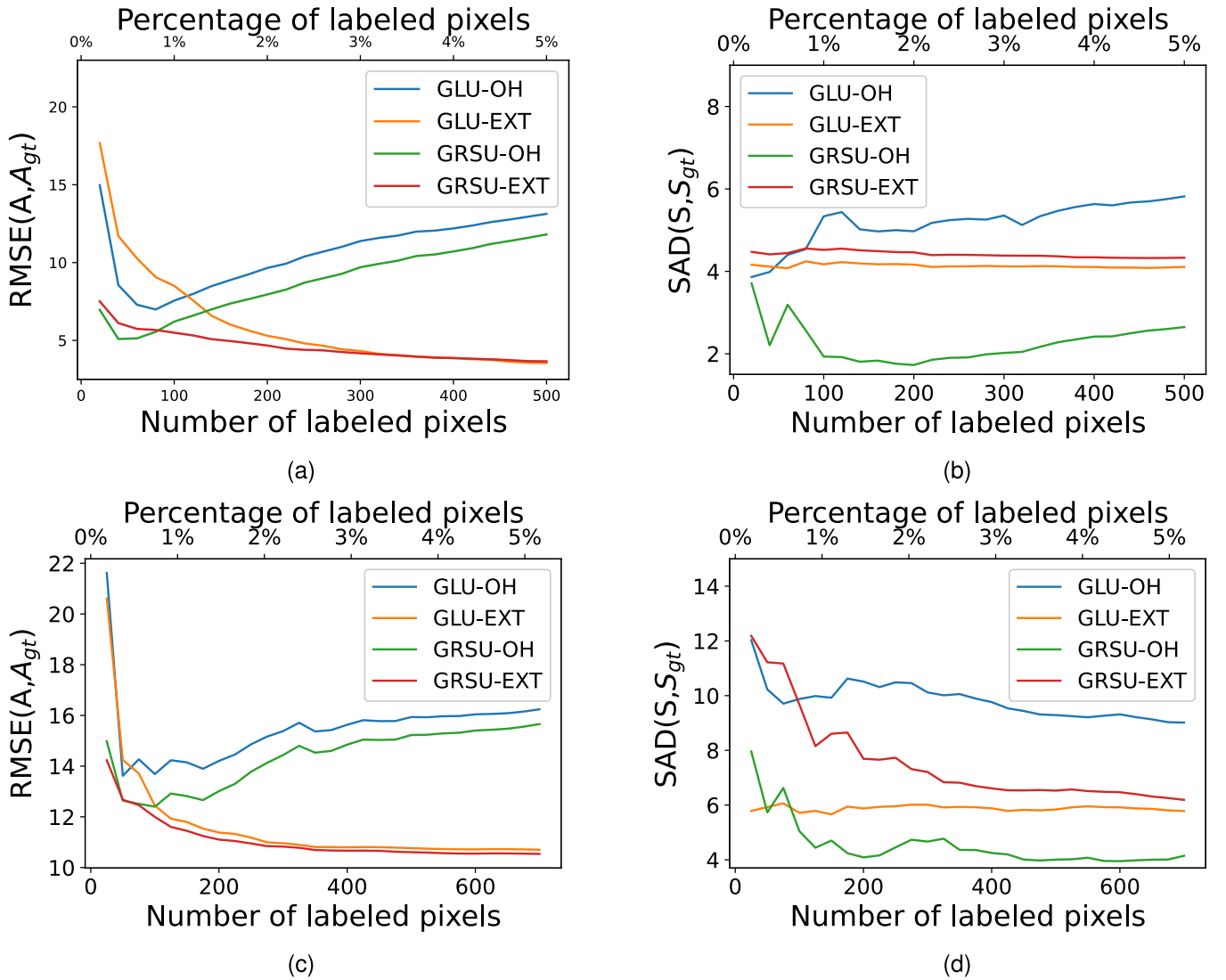


Fig. 10. RMSE and SAD curves with respect to the number of labeled pixels for the Jasper Ridge and the Apex datasets. For each plot, the x -axis on the top shows the percentage of labeled pixels, while the bottom one is the number of labeled pixels. In the active learning process, we apply the MCVOpt and VOpt acquisition functions for the Jasper Ridge and Apex datasets, respectively. Each curve starts with only one random pixel per end-member and sample up to 5% of labeled pixels. (a) Jasper RMSE. (b) Jasper SAD. (c) Apex RMSE. (d) Apex SAD.

λ , γ , and ρ are provided in Table II. The parameter α is designed to balance the numbers of the training pixels and total

pixels. Since we have various numbers of training pixels in this experimental setting, we choose $\alpha_0 = 10$ and $\alpha = C_\alpha \alpha_0 / N_{\text{train}}$,

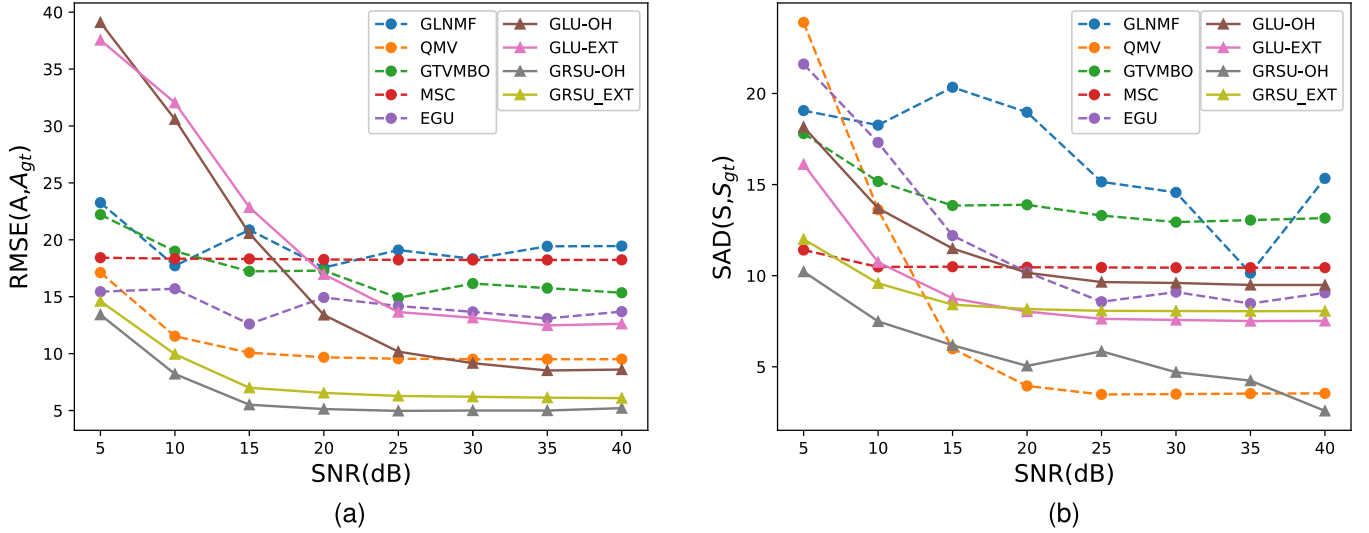


Fig. 11. RMSE (for A) and SAD (for S) curves with respect to the SNR values of noisy input of the Jasper Ridge dataset corrupted by Gaussian white noise. In both panels, the performance of other blind methods is illustrated by dashed lines, whereas our proposed semisupervised methods are represented by solid lines. (a) RMSE versus SNR. (b) SAD versus SNR.

where C_α is a constant depending on the dataset and N_{train} is the number of training pixels. In practice, $C_\alpha = 50$ for the Jasper Ridge dataset and $C_\alpha = 10$ for the Apex dataset.

Fig. 10 shows the changes in the RMSE on abundance and SAD on the end-members with respect to the number of labeled pixels obtained by active learning. Two abundance RMSE curves in Fig. 10(a) and (c) illustrate that our methods trained on OH pseudo-labels (GLU-OH and GRSU-OH) improve in the very beginning and deteriorate with more labeled pixels (after 1%), which is attributed to the inaccurate information of the OH labels. At the beginning of active learning, the most representative pixels are selected, whose ground-truth abundance vectors might be close to the OH vector. When we incorporate more OH labels for unmixing, the OH labels are misleading as opposed to the EXT abundance map. On the other hand, GLU-EXT and GRSU-EXT always benefit from increasing the number of labeled pixels increases, but with a diminishing gain after around 3% of the label rate. In practice, we would like to use the OH pseudo-labels for training, since it is much easier to obtain, but not exceeding 1% for the active learning process.

Compared with the RMSE curves, the end-member SAD curves are relatively more stable with respect to the increase in the number of training pixels. In practice, after 2% of the label rate, the SAD values do not change much for all our methods, which implies that only a small number of training pixels are needed to estimate the end-members.

Overall for both GLU and GRSU, the increase in the training pixels does not always deem improvements in the performances. In fact, including more OH pseudo-labels is not beneficial to unmixing. A label rate around 1% would be a good choice experimentally.

C. Robustness Study

In order to provide a quantitative validation of the robustness exhibited by our approach, we have undertaken an extensive investigation into the performances evaluated by RMSE (26)

and SAD (27) of various methods applied to the Jasper Ridge dataset. We add different amounts of Gaussian white noise to this data matrix, such that the resulting data have the signal-to-noise ratio (SNR) ranging from 5 to 40 dB with a 5-dB increment. Note that a higher SNR value implies less noisy data.

As evidenced in Fig. 11, the proposed GRSU method achieves the highest accuracy in estimating the abundance map A , no matter whether OH or EXT is considered under all the noise levels. The spectrum matrix S estimated by GRSU is the best for low SNR regime and is comparable to QMV when the data are less noisy ($\text{SNR} > 15$ dB). This result is reasonable, as we only impose two simple constraints on S (i.e., nonnegativity and sum-to-one). On the other hand, the GLU method seems sensitive to noise. In circumstances where the SNR is relatively low, the RMSE and SAD values produced by GLU tend to be large. But, we also observe a sharp decline in both RMSE and SAD by the GLU method with a gradual increase in SNR.

V. CONCLUSION

In this article, we propose two semisupervised hyperspectral models, the GLU and GRSU. GLU applies the graph Laplace learning directly to solve the HSU problem by regarding the class probability as the abundance map. Initialized by GLU, GRSU estimates the abundance map A and end-member spectrum matrix S by solving an energy-minimizing problem via an iterative ADMM scheme. We extended the graph Laplace learning to a regularized version and explored the close-form solutions for efficient computation. This regularized graph Laplace learning is a subproblem in the ADMM iteration process. We conducted extensive experiments using four standard hyperspectral datasets to compare our semisupervised methods with five state-of-the-art methods in hyperspectral blind unmixing. All the results demonstrated that the proposed GLU and GRSU methods have a significant improvement with a minimum amount of supervision. Furthermore, both methods can take either the ground-truth abundance maps or the OH

pseudo-labels as the training information, the latter of which is much easier to obtain, since it only requires determining the major end-member of each training pixel. According to our experiments, it is unnecessary to require the ground-truth abundance maps to feed in the semisupervised framework. Sometimes, models trained on pseudo-labels yield better performance than the EXT values. We also discussed the influence of the number of training pixels on the model performance, revealing that a label rate of 1% is experimentally sufficient for satisfactory results. In addition, our GRSU method has great robustness over Gaussian white noise. The new methods have a great potential for real-world problems, since they do not need a ground-truth abundance map and instead can work with pseudo-labels.

There are several promising directions following this work. First, scaling factor [1] and spectral variability [2] terms can be incorporated into our model to further enhance the unmixing performance. Second, Fig. 11(b) suggests a need for regularizations of the spectral matrix, in addition to nonnegativity and sum-to-one constraints, for improving the robustness of the proposed methods. Finally, the current graph Laplacian learning solver can be replaced with neural networks or graph neural network techniques.

APPENDIX

ACQUISITION FUNCTION FORMULAS

To make the paper self-contained, we provide the formulas for the VOpt [51] and MCVOpt [45] functions. Given a graph $G(\mathcal{X}, W)$, the formulas are based on a low-rank approximation of the graph Laplacian matrix L . With ordered eigenvalues of L as $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_N$, the truncated decomposition of L with the smallest $M < N$ eigenvalues has the form $L \approx V\Lambda V^\top$, where $\Lambda \in \mathbb{R}^{M \times M}$ is a diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \dots, \lambda_M$ and $V \in \mathbb{R}^{N \times M}$ is the matrix of corresponding eigenvectors. Furthermore, a Gaussian correlation matrix C is defined by

$$C = \left(\Lambda + V^\top \left(\frac{1}{\gamma^2} P^\top P \right) V \right)^{-1} \quad (28)$$

where $P \in \mathbb{R}^{|\mathcal{X}_L| \times N}$ is a projection matrix onto the indices corresponding to the label set $\hat{\mathcal{X}}$ and γ is a positive constant (we choose $\gamma = 0.1$ in this article). When the graph G is connected, such a matrix C exists, i.e., the matrix $\Lambda + V^\top ((1/\gamma^2) P^\top P) V$ is invertible.

At each iteration in the active learning process with the current label set $\hat{\mathcal{X}}$, graph Laplace learning gives the optimized node function \mathbf{u}^* (and corresponding node matrix U^*) by minimizing the energy (3). Denoting the k th column of V^\top by \mathbf{v}_k , the VOpt and MCVOpt acquisition functions are given by

$$\mathcal{A}_{\text{VOpt}}(\mathbf{x}_k) = \frac{1}{\gamma^2 + \mathbf{v}_k^\top C \mathbf{v}_k} \|C \mathbf{v}_k\|_2^2 \quad (29)$$

$$\mathcal{A}_{\text{MCVOpt}}(\mathbf{x}_k) = \frac{1}{\gamma^2 + \mathbf{v}_k^\top C \mathbf{v}_k} \|C \mathbf{v}_k\|_2^2 (U^*)^\top V \mathbf{v}_k - \mathbf{y}_k^* \|_2 \quad (30)$$

where \mathbf{y}_k^* is the OH thresholding of $\mathbf{u}^*(\mathbf{x}_k)$.

ACKNOWLEDGMENT

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Geospatial-Intelligence Agency.

REFERENCES

- [1] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, Aug. 2016.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
- [6] X.-R. Feng, H.-C. Li, R. Wang, Q. Du, X. Jia, and A. Plaza, "Hyperspectral unmixing based on nonnegative matrix factorization: A comprehensive review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4414–4436, 2022.
- [7] P. Sajda, S. Du, T. Brown, L. Parra, L. Parra, and R. Stoyanova, "Recovery of constituent spectra in 3D chemical shift imaging using non-negative matrix factorization," in *Proc. 4th Int. Symp. Independ. Compon. Anal. Blind Signal Separat.*, 2003, pp. 71–76.
- [8] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, Jul. 2006.
- [9] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Hoboken, NJ, USA: Wiley, 2009.
- [10] D. C. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
- [11] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [12] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4267–4279, Sep. 2016.
- [13] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.
- [14] L. Drumetz, T. R. Meyer, J. Chanussot, A. L. Bertozzi, and C. Jutten, "Hyperspectral image unmixing with endmember bundles and group sparsity inducing mixed norms," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3435–3450, Jul. 2019.
- [15] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.
- [16] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
- [17] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.
- [18] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2341–2357, Apr. 2019.

- [19] Y. Yuan, Z. Zhang, and Q. Wang, "Improved collaborative non-negative matrix factorization and total variation for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 998–1010, 2020.
- [20] L. Zhuang, C.-H. Lin, M. A. T. Figueiredo, and J. M. Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9858–9877, Dec. 2019.
- [21] A. L. Bertozzi and E. Merkurjev, "Graph-based optimization approaches for machine learning, uncertainty quantification and networks," in *Handbook of Numerical Analysis*, vol. 20. Amsterdam, The Netherlands: Elsevier, 2019, pp. 503–531.
- [22] K. Benzi, V. Kalofolias, X. Bresson, and P. Vandergheynst, "Song recommendation with non-negative matrix factorization and graph total variation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2439–2443.
- [23] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [24] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.
- [25] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 101–118, Feb. 2014.
- [26] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
- [27] M. Li, F. Zhu, A. J. X. Guo, and J. Chen, "A graph regularized multilinear mixing model for nonlinear hyperspectral unmixing," *Remote Sens.*, vol. 11, no. 19, p. 2188, Sep. 2019.
- [28] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the Nyström method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 214–225, Feb. 2004.
- [29] J. Qin, H. Lee, J. T. Chi, Y. Lou, J. Chanussot, and A. L. Bertozzi, "Fast blind hyperspectral unmixing based on graph Laplacian," in *Proc. 10th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Sep. 2019, pp. 1–5.
- [30] J. Qin et al., "Blind hyperspectral unmixing based on graph total variation regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3338–3351, Apr. 2021.
- [31] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *J. ACM*, vol. 45, no. 6, pp. 891–923, Nov. 1998.
- [32] B. Chen, K. Miller, A. L. Bertozzi, and J. Schwenk, "Graph-based active learning for surface water and sediment detection in multispectral images," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jun. 2023, pp. 1–4.
- [33] S. Ozkan, B. Kaya, and G. B. Akar, "EndNet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 482–496, Jan. 2019.
- [34] S. S. Vijayashankar, V. S. Deshpande, and J. S. Bhatt, "A practical approach for hyperspectral unmixing using deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [35] B. Rasti, B. Koirala, P. Scheunders, and J. Chanussot, "MiSiC-Net: Minimum simplex convolutional network for deep hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522815.
- [36] D. Hong et al., "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6518–6531, Nov. 2022.
- [37] Z. Han, D. Hong, L. Gao, J. Yao, B. Zhang, and J. Chanussot, "Multimodal hyperspectral unmixing: Insights from attention networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 25524913.
- [38] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, Jan. 2019.
- [39] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 35–49, Dec. 2019.
- [40] S. S. Vijayashankar, J. S. Bhatt, and B. Chattopadhyay, "Virtual dimensionality of hyperspectral data: Use of multiple hypothesis testing for controlling Type-I error," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2974–2985, 2020.
- [41] J. S. Bhatt and M. V. Joshi, "Deep learning in hyperspectral unmixing: A review," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 2189–2192.
- [42] B. Zhang et al., "Progress and challenges in intelligent remote sensing satellite systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1814–1822, 2022.
- [43] K. Miller, H. Li, and A. L. Bertozzi, "Efficient graph-based active learning with Probit likelihood via Gaussian approximations," in *Proc. Workshop Real World Exp. Design Act. Learn.*, Jul. 2020, pp. 1–8.
- [44] K. Miller and A. L. Bertozzi, "Model-change active learning in graph-based semi-supervised learning," 2021, *arXiv:2110.07739*.
- [45] K. Miller et al., "Graph-based active learning for semi-supervised classification of SAR data," *Proc. SPIE*, vol. 12095, May 2022, Art. no. 120950C.
- [46] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [47] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.
- [48] 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.
- [49] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020.
- [50] B. Chen et al., "Batch active learning for multispectral and hyperspectral image segmentation using similarity graphs," *Commun. Appl. Math. Comput.*, 2023, doi: [10.1007/s42967-023-00284-8](https://doi.org/10.1007/s42967-023-00284-8).
- [51] M. Ji and J. Han, "A variance minimization criterion to active learning on graphs," in *Proc. 5th Int. Conf. Artif. Intell. Statist.*, 2012, pp. 556–564.
- [52] Y. Qiao et al., "Uncertainty quantification for semi-supervised multi-class classification in image processing and ego-motion analysis of body-worn videos," *Electron. Imag.*, vol. 31, no. 11, pp. 264–1–264–7, Jan. 2019.
- [53] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, no. 11, pp. 2399–2434, 2006.
- [54] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.
- [55] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. 20th Int. Conf. Mach. Learn.* Palo Alto, CA, USA: AAAI Press, 2003, pp. 912–919.
- [56] J. Chapman, B. Chen, Z. Tan, J. Calder, K. Miller, and A. L. Bertozzi, "Novel batch active learning approach and its application to synthetic aperture radar datasets," *Proc. SPIE*, vol. 12520, Jun. 2023, Art. no. 125200B.
- [57] W. Wang and M. Á. Carreira-Perpiñán, "Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application," 2013, *arXiv:1309.1541*.
- [58] L. C. Evans, *Partial Differential Equations*. Providence, RI, USA: American Mathematical Society, 1998.
- [59] B. Ousting, C. D. White, and É. Oudet, "Minimal Dirichlet energy partitions for graphs," *SIAM J. Sci. Comput.*, vol. 36, no. 4, pp. 1635–1651, Jan. 2014.
- [60] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, nos. 1–3, pp. 293–318, Apr. 1992.
- [61] M. Hasanlou, *Remote Sensing Datasets*. Accessed: Mar. 10, 2023. [Online]. Available: <https://rslab.ut.ac.ir/data>

- [62] M. E. Schaepman et al., "Advanced radiometry measurements and Earth science applications with the airborne prism experiment (APEX)," *Remote Sens. Environ.*, vol. 158, pp. 207–219, Mar. 2015.
- [63] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.



Bohan Chen (Student Member, IEEE) received the B.S. degree in mathematics from the School of Mathematical Sciences, Peking University, Beijing, China, in 2018. He is currently pursuing the Ph.D. degree in mathematics with the University of California at Los Angeles (UCLA), Los Angeles, CA, USA.

He conducted his research at both UCLA and the Los Alamos National Laboratory (LANL), Los Alamos, NM, USA. His research interests include optimal control problems, image processing,

graph semisupervised learning, and active learning.

Mr. Chen received the UC-National Laboratory In-Residence Graduate Fellowship in 2021.



Yifei Lou received the Ph.D. degree in applied mathematics from the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, in 2010.

She was a Post-Doctoral Fellow at the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, followed by another post-doctoral training at the Department of Mathematics, University of California at Irvine, Irvine, CA, USA, from 2012 to 2014. She was on the faculty with the Mathematical Sciences Department,

The University of Texas at Dallas, Richardson, TX, USA, from 2014 to 2023, first as an Assistant Professor and then as an Associate Professor. She holds a joint position with the Department of Mathematics and the School of Data Science and Society (SDSS), The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. Her research interests include compressive sensing and its applications, image analysis (medical, hyperspectral, and seismic imaging), and (nonconvex) optimization algorithms.

Dr. Lou received the National Science Foundation CAREER Award in 2019.



Andrea L. Bertozzi (Member, IEEE) received the A.B., M.A., and Ph.D. degrees from Princeton University, Princeton, NJ, USA, in 1987, 1988, and 1991, respectively.

She was an L.E. Dickson Instructor and an NSF Post-Doctoral Fellow at The University of Chicago, Chicago, IL, USA, from 1991 to 1995. She was the Maria Geoppert-Mayer Distinguished Scholar with the Argonne National Laboratory, Lemont, IL, USA, from 1995 to 1996. She was on the faculty at Duke University, Durham, NC, USA, from 1995 to 2004,

first as an Associate Professor of mathematics and then as a Professor of mathematics and physics. Since 2003, she has been on the faculty at the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, where she is currently a Distinguished Professor of mathematics and mechanical and aerospace engineering, along with the Betsy Wood Knapp Chair for innovation and creativity. She is an Applied Mathematician with expertise in nonlinear partial differential equations and graphical models for machine learning.

Dr. Bertozzi was elected to the American Academy of Arts and Sciences in 2010 and to the fellows of the Society of Industrial and Applied Mathematics (SIAM) in 2010. She became a fellow of the American Mathematical Society in 2013 and the American Physical Society in 2016. In 2018, she was elected to the U.S. National Academy of Sciences. Her honors include the Sloan Research Fellowship in 1995, the Presidential Early Career Award for Scientists and Engineers in 1996, the SIAM Kovalevsky Prize in 2009, and the SIAM Kleinman Prize in 2019. She received the SIAM Outstanding Paper Prize in 2014 with A. Flenner, for her work on geometric graph-based algorithms for machine learning. She received the Simons Math + X Investigator Award in 2017.



Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

From 2015 to 2017, he was a Visiting Professor at the University of California at Los Angeles (UCLA), Los Angeles, CA, USA. Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. He has been a

Visiting Scholar at Stanford University, Stanford, CA, USA; the KTH Royal Institute of Technology, Stockholm, Sweden; and the National University of Singapore (NUS), Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland. He holds the AXA Chair in remote sensing and is an Adjunct Professor at the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Dr. Chanussot was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008 and the Institut Universitaire de France from 2012 to 2017. He is a fellow of European Laboratory for Learning and Intelligent Systems (ELLIS) and the Asia-Pacific Artificial Intelligence Association. He is the Founding President of the IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010, which received the 2010 IEEE Geoscience and Remote Sensing Society (GRSS) Chapter Excellence Award. He has received multiple outstanding paper awards. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He was the Chair from 2009 to 2011 and Cochair from 2005 to 2008 of the GRS Data Fusion Technical Committee. He was the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. From 2017 to 2019, he was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia. He has been a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters) since 2018. He was an Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. In 2014, he served as a Guest Editor for *IEEE Signal Processing Magazine*. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and the PROCEEDINGS OF THE IEEE.