

## Multiple instance hybrid estimator for hyperspectral target characterization and sub-pixel target detection

Changzhe Jiao<sup>a,\*</sup>, Chao Chen<sup>b</sup>, Ronald G. McGarvey<sup>c</sup>, Stephanie Bohlman<sup>d</sup>, Licheng Jiao<sup>a</sup>, Alina Zare<sup>e,\*</sup>

<sup>a</sup> Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education of China, School of Artificial Intelligence, Xidian University, Xi'an 710071, China

<sup>b</sup> MathWorks, Natick, MA 01760, USA

<sup>c</sup> Department of Industrial and Manufacturing Systems Engineering, University of Missouri, Columbia, MO 65211, USA

<sup>d</sup> School of Forest Resources and Conservation, University of Florida, Gainesville, FL 32611, USA

<sup>e</sup> Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA



### ARTICLE INFO

#### Keywords:

Target detection  
Hyperspectral  
Endmember extraction  
Multiple instance learning  
Hybrid detector  
Target characterization

### ABSTRACT

The Multiple Instance Hybrid Estimator for discriminative target characterization from imprecisely labeled hyperspectral data is presented. In many hyperspectral target detection problems, acquiring accurately labeled training data is difficult. Furthermore, each pixel containing target is likely to be a mixture of both target and non-target signatures (*i.e.*, sub-pixel targets), making extracting a pure prototype signature for the target class from the data extremely difficult. The proposed approach addresses these problems by introducing a data mixing model and optimizing the response of the hybrid sub-pixel detector within a multiple instance learning framework. The proposed approach iterates between estimating a set of discriminative target and non-target signatures and solving a sparse unmixing problem. After learning target signatures, a signature based detector can then be applied on test data. Both simulated and real hyperspectral target detection experiments show the proposed algorithm is effective at learning discriminative target signatures and achieves superior performance over state-of-the-art comparison algorithms.

### 1. Introduction

Hyperspectral imaging spectrometers collect electromagnetic energy scattered in the scene across hundreds or thousands of spectral bands, capturing both the spatial and spectral information (Landgrebe, 2002). The spectral information is a combination of the reflection and/or emission of sunlight across wavelength by objects on the ground, and contains the unique spectral characteristics of different materials (Keshava and Mustard, 2002; Bioucas-Dias et al., 2012). The wealth of spectral information in hyperspectral imagery enables the possibility to conduct sub-pixel analysis in application areas including target detection (Yuksel et al., 2015; Zare et al., 2008), precision agriculture (Mahajan et al., 2014; Wang et al., 2012), biomedical applications (Pike et al., 2016; Pardo et al., 2017) and others (Eismann et al., 2009; Lara et al., 2013).

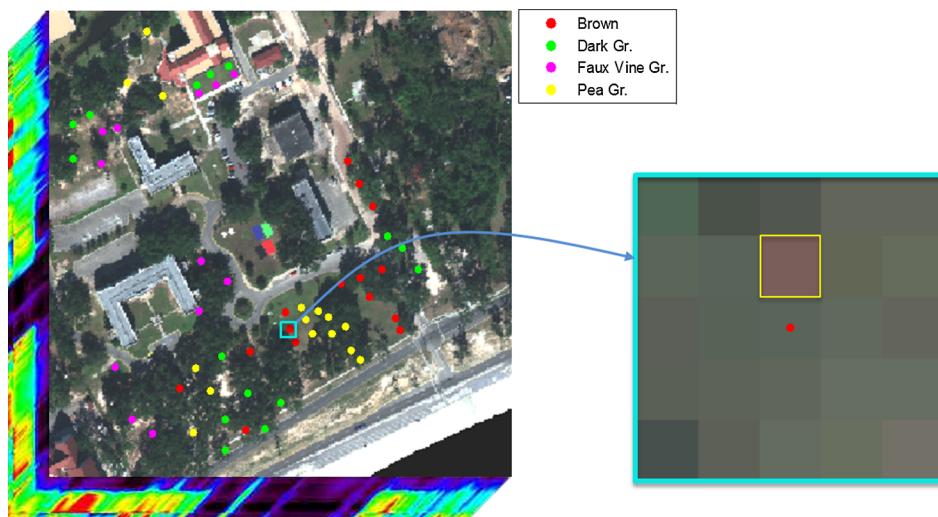
Hyperspectral target detection generally refers to the task of locating all instances of a target given a known spectral signature within a hyperspectral scene. The reasons many classification methods are not

applicable to hyperspectral target detection tasks are threefold:

1. Precise training labels for sub-pixel targets are often difficult or infeasible to obtain. For example, the ground truth information is often obtained using a Global Positioning System (GPS) receiver placed at the target location. However, the accuracy of the GPS coordinates could drift for several meters depending on the accuracy of the GPS system.
2. The number of training instances from the positive (target) class is often small compared to that of the negative training data such that training an effective classifier is difficult. A hyperspectral image with hundreds of thousands of pixels may only have a few pixels or sub-pixel level target points.
3. Due to the relatively low spatial resolution of hyperspectral imagery and the diversity of natural scenes, many targets are mixed points (sub-pixel targets) and the amount of the target proportions are not known.

\* Corresponding authors.

E-mail addresses: [cjiao@xidian.edu.cn](mailto:cjiao@xidian.edu.cn) (C. Jiao), [azare@ece.ufl.edu](mailto:azare@ece.ufl.edu) (A. Zare).



**Fig. 1.** Illustration of inaccurate coordinates from GPS: one target denoted as brown by GPS has one pixel drift.

For the above reasons, signature based hyperspectral target detection (Manolakis and Shaw, 2002; Nasrabadi, 2014) is generally applied over classification approaches. However, the performance of signature based detectors relies heavily on the quality of the target spectral signature used. Yet, obtaining an effective target signature can be a challenging problem. Sub-pixel targets are difficult to see in the imagery and the co-registration of the targets to GPS coordinates could drift across several pixels.

As an example, the left image of Fig. 1 shows the scattered target locations over the MUUFL Gulfport data set collected over the University of Southern Mississippi-Gulfpark Campus (Gader and Zare, 2013), where there are 4 types of targets throughout the scene: Brown (15 examples), Dark Green (15 examples), Faux Vineyard Green (12 examples) and Pea Green (15 examples). The region highlighted by blue<sup>1</sup> rectangle shown in the left image of Fig. 1 contains one brown target. The image in the right of Fig. 1 shows a zoomed in view of this region. From this zoomed image, we can clearly see that the GPS coordinate of this brown target (denoted as red dot) is shifted one pixel from the actual brown target location (denoted as yellow rectangle). This is a rare example where we can visually see the brown target. Most of the targets are difficult to distinguish visibly since they are highly mixed and sub-pixel.

In this paper, we model the hyperspectral target estimation task as a multiple instance concept learning problem (Dietterich et al., 1997; Maron and Lozano-Perez, 1998) and present the multiple instance hybrid estimator (MI-HE) for characterization of a target signature from imprecisely labeled hyperspectral imagery. Here, *concepts* refer to the generalized class prototypes in the feature space. In the case of hyperspectral image unmixing, a concept is the spectral signature of one assumed pure material in the scene, also known as an endmember. MI-HE explicitly addresses the aforementioned problems from the following aspects:

1. **Uncertain labels.** MI-HE adopts the idea of “bags” from multiple instance learning (MIL), introduces the multiple instance data mixing model and proposes a multiple instance concept learning framework to address the uncertain labels in hyperspectral target detection, e.g., target locations coming from GPS or region of interest manually denoted by analysts.
2. **Unbalanced number of target and non-target pixels in training data.** MI-HE addresses this problem by applying a signature-based

detector only to the pixels from possible target regions denoted by a GPS in the training step. The non-target regions are only needed to refine the background concepts.

3. **Mixed training data.** MI-HE models each pixel as a linear combination of target and/or non-target concepts, so the estimated target and background concepts applied to testing data can be used to perform sub-pixel detection.

Several previous methods for target characterization have been developed in the literature. The FUnctions of Multiple Instance (FUMI) algorithms (Jiao and Zare, 2015; Jiao and Zare, 2016) learn representative concept from reconstruction error of the uncertainly labeled data. Compared with FUMI, MI-HE learns a discriminative target concept that maximizes the detection response of possible target regions so the estimated signatures are more discriminative for target detection. Furthermore, the FUMI algorithms do not exploit the complete label information from the training data, i.e., the FUMI algorithms combine all possible target regions together into one target bag and thus discard the information that each target region contains at least one target instance. Thus is not applicable to some target detection problems, e.g., negative bags cannot provide the entire background information. On the contrary, MI-HE adopts a generalized mean model to differentiate each individual target region. Compared to the discriminative target characterization algorithms, MI-ACE and MI-SMF (Zare et al., 2017) (which maximize the matched filter response in an MIL framework and estimates only one target signature), MI-HE has the potential to learn multiple signatures to address signature variability.

Expanding upon our work in Jiao and Zare (2017), the MI-HE algorithm and experiments presented in this paper maintain the following improvements and advantages in comparison to our prior work: (1) introducing a discriminative concept learning term; (2) improved gradient descent optimization using Armijo's rule; (3) comprehensive experiments on multiple concepts learning (both simulated and realistic) and more types of targets for Gulfport data; (4) comprehensive comparison with state-of-the-art MIL algorithms; (5) analysis of MI-HE's robustness to parameter setting.

## 2. Related methods

Multiple instance learning (MIL) was first investigated by Dietterich et al. (1997) in the 1990s for the prediction of drug activity (musk activity). In MIL, training data is partitioned into sets of labeled “bags” (instead of being individually labeled) in which a bag is defined to be a multi-set of data points. A positive bag must contain at least one true

<sup>1</sup> For interpretation of color in Fig. 1, the reader is referred to the web version of this article.

positive (target) data point and negative bags are composed entirely of negative data. Thus, data point-specific training labels are unavailable. Given training data in this form, the majority of MIL methods either: (1) learn target concepts for describing the target class; or (2) train a classifier that can distinguish between individual target and non-target data points and/or bags.

Since the introduction of the MIL framework (Dietterich et al., 1997), many methods have been proposed and developed in the literature. The majority of MIL approaches focus on learning a classification decision boundary to distinguish between positive and negative instances/bags from the ambiguously labeled data. The mi-SVM (Andrews et al., 2002) models the MIL problem as a generalized mixed integer formulation of support vector machine and was solved iteratively between training a regular SVM and heuristic reassignment of the training labels. The Multiple-Instance Learning via Embedded Instance Selection (MILES) (Chen et al., 2006) relaxes the constraint in MIL that negative bags are composed of all negative instances and allows target concept to be related to negative bags for a more general application in computer vision. MILES proposes to first embed each bag to a target concept based feature space, where the set of candidate target concept comes from the union of all bags. Then a 1-norm SVM (Zhu et al., 2004) is trained on the feature vectors extracted from each bag; finally, an instance selection is performed based on the SVM decision value to realize instance-level classification.

Although the above mentioned approaches are effective for training classifiers given imprecise labels, they generally do not provide an intuitive description or *representative concept* that characterizes the salient and discriminative features of the target class. The few existing approaches that estimate a target concept include Diverse Density (DD) (Maron and Lozano-Perez, 1998) that estimates a concept by minimizing its distance to at least one instance from each positive bag and maximizing its distance from all instances in negative bags. The Expectation-Maximization (EM) version of diverse density (EM-DD) (Zhang and Goldman, 2002) iteratively estimates which instances in each positive bag belong to the target class and then only uses those points from the positive bags to estimate a target concept that maximizes the diverse density. eFUMI (Zare and Jiao, 2014; Jiao and Zare, 2015) treats each instance as a convex combination of positive and/or negative concepts and estimates the target and non-target concepts using an EM approach. The MI-SMF and MI-ACE (Zare et al., 2017) maximize the response of SMF (spectral matched filter (Theiler and Foy, 2006; Nasrabadi, 2008)) and ACE (adaptive cosine/coherent estimator (Kraut and Scharf, 1999; Kraut et al., 2001)) respectively under a multiple instance learning framework and efficiently learn a discriminative target signature. However, most of these prototype-based methods only find a single target concept and are, thus, unable to account for large variation in the target class.

### 3. Multiple instance hybrid estimator

Let  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$  be training data where  $d$  is the dimensionality of an instance and  $N$  is the total number of training instances. The data are grouped and arranged into  $K$  bags,  $\mathbf{B} = \{\mathbf{B}_1, \dots, \mathbf{B}_K\}$ , such that the first  $K^+$  bags are positively labeled with associated binary bag-level label  $L_i = 1$ ,  $i = 1, \dots, K^+$ , and the rest  $K^-$  bags are negatively labeled with associated binary bag-level label  $L_i = 0$ ,  $i = K^+ + 1, \dots, K$ .  $N^+$  and  $N^-$  are the total number of positive instances and negative instances, as indicated below:

$$N = N^+ + N^- = \sum_{i=1}^{K^+} N_i + \sum_{i=K^++1}^K N_i, \quad (1)$$

where  $N_i$  is the number of instances in bag  $\mathbf{B}_i$ .  $\mathbf{x}_{ij} \in \mathbf{B}_i$  denotes the  $j^{th}$  instance in bag  $\mathbf{B}_i$  with instance-level label  $l_{ij} \in \{0, 1\}$ .

Since the MIL problem states that there must be at least one positive instance in each positive bag and each negative bag must consist of only

negative instances, we can approximate the probability of an individual bag to the instances in each bag, as shown in Eq. (2). Specifically, the probability for a positive bag to be positive is substituted by the instance in this bag with highest “positiveness” and the probability for a negative bag to be negative is represented by the joint probability of all instances in this bag to be negative.

$$J_1 = \prod_{i=1}^{K^+} \max_{\mathbf{x}_{ij} \in \mathbf{B}_i} Pr(l_{ij} = +|\mathbf{B}_i) \prod_{i=K^++1}^K \prod_{j=1}^{N_i} Pr(l_{ij} = -|\mathbf{B}_i). \quad (2)$$

Eq. (2) contains a max operation that is difficult to optimize numerically. Some algorithms in the literature (Maron and Ratan, 1998; Zhang and Goldman, 2002) adopt a noisy-OR model instead of using max. However, experimental results show that the noisy-OR model is highly non-smooth and needs to be repeated with many different initializations (typically using every positive training instance) to avoid local optima. In order to find some alternatives for max, a lot of effort has been spent. Babenko et al. (2008) summarized four methods for transferring the max operation to differentiable a ‘softmax’. The generalized mean model (Bullen et al., 1988) is one such ‘softmax’ solution. The generalized mean has been applied as an alternative to max in many MIL problems including cancer prediction (Xu et al., 2012; Quellec et al., 2016), image segmentation (Wu et al., 2014; Kraus et al., 2016) and face recognition (Wohlhart et al., 2011). In our proposed approach, we adopt the generalized mean as an alternative to max operation in Eq. (2), as shown in Eq. (3).

$$J_2 = \prod_{i=1}^{K^+} \left( \frac{1}{N_i} \sum_{j=1}^{N_i} Pr(l_{ij} = +|\mathbf{B}_i)^b \right)^{\frac{1}{b}} \prod_{i=K^++1}^K \prod_{j=1}^{N_i} Pr(l_{ij} = -|\mathbf{B}_i), \quad (3)$$

where  $b \in [-\infty, +\infty]$  is a real number controlling the function to approximately vary from min to max.

Then taking the negative logarithm and scaling the second term of Eq. (3) result in:

$$\begin{aligned} -\ln J_2 = & - \sum_{i=1}^{K^+} \frac{1}{b} \ln \left( \frac{1}{N_i} \sum_{j=1}^{N_i} Pr(l_{ij} = +|\mathbf{B}_i)^b \right) \\ & - \rho \sum_{i=K^++1}^K \sum_{j=1}^{N_i} \ln Pr(l_{ij} = -|\mathbf{B}_i), \end{aligned} \quad (4)$$

where the scaling factor  $\rho$  is usually set to be smaller than one to control the influence of negative bags.

Here each instance is modeled as a sparse linear combination of target and/or background concepts  $\mathbf{D}, \mathbf{x} \approx \mathbf{D}\mathbf{a}$ , where  $\mathbf{D} = [\mathbf{D}^+ \ \mathbf{D}^-] \in \mathbb{R}^{d \times (T+M)}$ ,  $\mathbf{D}^+ = [\mathbf{d}_1, \dots, \mathbf{d}_T]$  is the set of  $T$  target concepts and  $\mathbf{D}^- = [\mathbf{d}_{T+1}, \dots, \mathbf{d}_{T+M}]$  is the set of  $M$  background concepts,  $\mathbf{a}$  is the sparse vector of weights for instance  $\mathbf{x}$ . Solving  $\mathbf{a}$  for data  $\mathbf{x}$  given dictionary set  $\mathbf{D}$  is modeled as a Lasso problem (Tibshirani, 1996; Chen et al., 2001) shown in Eq. (5):

$$\mathbf{a}^* = \operatorname{argmin} \frac{1}{2} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1, \quad (5)$$

where  $\lambda$  is a regularization constant to control the sparsity of  $\mathbf{a}$ . The solving of  $l_1$  regularized least squares have been investigated extensively in the literature (Mallat, 2008; Bach et al., 2012; Mairal et al., 2014). Here we adopt the iterative shrinkage-thresholding algorithm (ISTA) (Figueiredo and Nowak, 2003; Daubechies et al., 2004) for solving the sparse codes  $\mathbf{a}$ .

Since in MIL the positive bags are mixture of both true positive points and false positive points, the hybrid detector (Broadwater et al., 2004; Broadwater and Chellappa, 2007) was introduced to the proposed objective function (4) to determine if instances from positive bags are the true positive points. Specifically, define the following detection statistic as an approximation of  $Pr(l_{ij} = +|\mathbf{B}_i)$ ,

$$\Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i, L_i = 1) = \exp\left(-\beta \frac{\|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^-\mathbf{p}_{ij}\|^2}\right), \quad (6)$$

where  $\beta$  is a scaling parameter,  $\mathbf{a}_{ij}$  and  $\mathbf{p}_{ij}$  are the sparse representation of  $\mathbf{x}_{ij}$  given the entire concept set  $\mathbf{D}$  and background concept set  $\mathbf{D}^-$ . Further,  $\mathbf{a}_{ij} = [\mathbf{a}_{ij}^+; \mathbf{a}_{ij}^-]$ , where  $\mathbf{a}_{ij}^+$  and  $\mathbf{a}_{ij}^-$  are sub-vectors of  $\mathbf{a}_{ij}$  corresponding to  $\mathbf{D}^+$  and  $\mathbf{D}^-$ , respectively. Since  $\mathbf{D}$  is a super set of  $\mathbf{D}^-$ , theoretically the reconstruction error of  $\mathbf{x}_{ij}$  using  $\mathbf{D}$  (the numerator) should not be greater than that using  $\mathbf{D}^-$  (the denominator). The residual vectors,

$$\begin{cases} \mathbf{r}_{ij} = (\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}) \\ \mathbf{q}_{ij} = (\mathbf{x}_{ij} - \mathbf{D}^-\mathbf{p}_{ij}), \end{cases} \quad (7)$$

are the reconstruction residual vectors corresponding to  $\mathbf{D}$  and  $\mathbf{D}^-$ , respectively.

The hybrid detector  $\Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i, L_i = 1)$  defined in Eq. (6) indicates that if a point  $\mathbf{x}_{ij} \in \mathbf{B}_i, L_i = 1$ , is a false positive point ( $l_{ij} = 0$ ) in a positively labeled bag, it should be well represented by both the non-target concepts  $\mathbf{D}^-$  and the entire concepts  $\mathbf{D}$ , respectively. Since this false positive point is a none-target point, the residual errors by both reconstruction methods should be of the same magnitude, i.e.,  $\|\mathbf{r}_{ij}\|^2 \approx \|\mathbf{q}_{ij}\|^2$ . After being scaled by  $\beta$ , the detection statistic of the hybrid detector for this false positive point,  $\Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i, L_i = 1) = \exp\left(-\beta \frac{\|\mathbf{r}_{ij}\|^2}{\|\mathbf{q}_{ij}\|^2}\right) \rightarrow 0$ . However, if  $\mathbf{x}_{ij} \in \mathbf{B}_i, L_i = 1$ , is a true positive point ( $l_{ij} = 1$ ) in the positively labeled bag, it should not be well represented by only the non-target concepts, so the residual error approximated by the entire concepts,  $\|\mathbf{r}_{ij}\|^2$ , will be much smaller than that by the background concepts,  $\|\mathbf{q}_{ij}\|^2$ , thus the detection for this true positive point,  $\Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i, L_i = 1) = \exp\left(-\beta \frac{\|\mathbf{r}_{ij}\|^2}{\|\mathbf{q}_{ij}\|^2}\right) \rightarrow 1$ . Here, the scaling factor  $\beta$  for this case is no longer consequential since  $\|\mathbf{r}_{ij}\|^2 \ll \|\mathbf{q}_{ij}\|^2$ .

For points from negative bags, the negative logarithm of  $Pr(l_{ij} = -|\mathbf{B}_i, L_i = 0)$  is modeled as the least squares of  $\mathbf{x}_{ij}$  shown in Eq. (8),

$$-\ln Pr(l_{ij} = -|\mathbf{B}_i, L_i = 0) = \|\mathbf{x}_{ij} - \mathbf{D}^-\mathbf{p}_{ij}\|^2, \quad (8)$$

where  $\mathbf{p}_{ij}$  is the sparse representation of  $\mathbf{x}_{ij}$  given  $\mathbf{D}^-$  and is solved by Eq. (5). Here instead of applying the hybrid detector, we use least squares to represent the residual error of  $\mathbf{x}_{ij}$ . This indicates that the negative points should be fully represented by only the non-target concepts,  $\mathbf{D}^-$ . The intuitive understanding of this assumption is that minimizing the least squares of all of the negative points provides a good description of the background. Moreover, because there are typically many more negative points in negatively labeled bags than true positive points in positive bags, the target concept estimation may be biased if the hybrid detector was also applied to negative instances.

Thus far, we have defined the probability terms in objective function 4 which aims to learn a set of concepts that maximize the hybrid sub-pixel detector statistic of the positive bags and characterize the negative bags. However, given the objective function so far, there is no guarantee that the estimated target concept captures only the distinct features of the positive class and is discriminative from the negative class. Inspired by the discriminative terms proposed by the Dictionary Learning with Structured Incoherence (Ramirez et al., 2010) and the Fisher Discrimination Dictionary Learning (FDDL) algorithm (Yang et al., 2011; Yang et al., 2014), we propose a cross incoherence term  $Q(\mathcal{X}, \mathbf{D}^+, \mathcal{A})$  shown in Eq. (9) to complete the objective, where  $\mathcal{X}$  is the concatenation of all instances from negatively labeled bags,  $\mathbf{D}^+$  is the target concept set which is the subset of  $\mathbf{D} = [\mathbf{D}^+ \mathbf{D}^-]$ ,  $\mathcal{A} = [\mathcal{A}^+ \mathcal{A}^-]$  is the sparse codes matrix of  $\mathcal{X}$  with respect to the entire concepts  $\mathbf{D}$ .

$$\begin{aligned} Q(\mathcal{X}, \mathbf{D}^+, \mathcal{A}) &= \frac{\alpha}{2} \|Diag((\mathbf{D}^+\mathcal{A}^+)^T \mathcal{X})\|_2^2 \\ &= \frac{\alpha}{2} \sum_{i=K^++1}^K \sum_{j=1}^{N_i} ((\mathbf{D}^+\mathbf{a}_{ij}^+)^T \mathbf{x}_{ij})^2 \end{aligned} \quad (9)$$

The understanding of the proposed cross incoherence term is presented by examining the reconstruction of the negative data set  $\mathcal{X}$ . First of all,  $\mathcal{X}$  should be well represented by the non-target concept set  $\mathbf{D}^-$ , i.e.,  $\mathcal{X} \approx \mathbf{D}^-\mathbf{P}$ . This is fulfilled by inclusion of the term in Eq. (8). Second, since  $\mathbf{D} = [\mathbf{D}^+ \mathbf{D}^-]$  is a superset of  $\mathbf{D}^-$ , the reconstruction error of  $\mathcal{X}$  by the entire concept set  $\mathbf{D}$  is also small, i.e.,  $\mathcal{X} \approx \mathbf{D}^+\mathcal{A}^+ + \mathbf{D}^-\mathcal{A}^- = \mathbf{R}^+ + \mathbf{R}^-$ . In order to have a target concept  $\mathbf{D}^+$  that is distinct from the negative data, it is expected that the reconstruction of  $\mathcal{X}$  with respect to the target concept,  $\mathbf{R}^+$ , should either maintain small energy or else have a bad representation of  $\mathcal{X}$ , and thus Eq. (9) is optimized.

The final objective function is shown in Eq. (10), which contains three terms: generalized mean (GM) term (first), background data fidelity term (second) and the cross incoherence (discriminative) term (third):

$$\begin{aligned} J_3 = & \sum_{i=1}^{K^+} \frac{1}{b} \ln \left( \frac{1}{N_i} \sum_{j=1}^{N_i} \exp\left(-\beta \frac{\|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^-\mathbf{p}_{ij}\|^2}\right)^b \right) \\ & + \rho \sum_{i=K^++1}^K \sum_{j=1}^{N_i} \|\mathbf{x}_{ij} - \mathbf{D}^-\mathbf{p}_{ij}\|^2 \\ & + \frac{\alpha}{2} \sum_{i=K^++1}^K \sum_{j=1}^{N_i} ((\mathbf{D}^+\mathbf{a}_{ij}^+)^T \mathbf{x}_{ij})^2, \end{aligned} \quad (10)$$

#### 4. Optimization

The optimization of Eq. (10) can be decomposed into two sub-problems, updating the concepts  $\mathbf{D}$  and the sparse representation  $\mathbf{a}$  alternatively. The algorithm stops when the number of preset iterations is reached or the change between two iterations is smaller than a preset threshold. The method is summarized in Algorithm 1 (Jiao and Zare, 2018). For readability, the derivation of update equations are described in Appendix A.

##### Algorithm 1. MI-HE algorithm

```

Input: MIL training bags  $\mathbf{B} = \{\mathbf{B}_1, \dots, \mathbf{B}_K\}$ , MI-HE parameters
1: Initialize  $\mathbf{D}^0, iter = 0$ 
2: repeat
3:   for  $t = 1, \dots, T$  do
4:     Solve  $\mathbf{a}_{ij}, \mathbf{p}_{ij}$  according to (A.16),
    $\forall i \in \{1, \dots, K\}, j \in \{1, \dots, N_i\}$ 
5:     Update  $\mathbf{d}_t$  using gradient descent according to (A.9)
6:      $\mathbf{d}_t \leftarrow \frac{1}{\|\mathbf{d}_t\|_2} \mathbf{d}_t$ 
7:   end for
8:   for  $k = T + 1, \dots, T + M$  do
9:     Solve  $\mathbf{a}_{ij}, \mathbf{p}_{ij}$  according to (A.16),
    $\forall i \in \{1, \dots, K\}, j \in \{1, \dots, N_i\}$ 
10:    Update  $\mathbf{d}_k$  using gradient descent according to (A.10)
11:     $\mathbf{d}_k \leftarrow \frac{1}{\|\mathbf{d}_k\|_2} \mathbf{d}_k$ 
12:  end for
13:   $iter \leftarrow iter + 1$ 
14: until Stopping criterion reached
15: return  $\mathbf{D}$ 

```

The initialization of target concepts in  $\mathbf{D}$  is conducted by computing the mean of  $T$  random subsets drawn from the union of all positive training bags. VCA (vertex component analysis) (Nascimento and Dias,

2005)) was applied to the union of all negative bags and the  $M$  vertices) were set as the initial background concepts.

## 5. Experimental results

In the following, MI-HE is evaluated and compared to several MIL concept learning methods on simulated data and to a real hyperspectral target detection data set. The simulated data experiments are included to illustrate the properties of MI-HE and provide insight into how and when the methods are effective.

For the experiments conducted in this paper, the parameter settings of the proposed MI-HE and comparison algorithms were optimized using a grid search on the first task of each experiment and then applied to the remaining tasks. For example, for mi-SVM classifier on the Gulfport Brown target task, the  $\gamma$  value of the RBF kernel was firstly varied from 0.5 to 5 at a step size of 0.5, and then a finer search around the current best value (with the highest AUC) at a step of 0.1 was performed. For algorithms with stochastic result, e.g., EM-DD, eFUMI, each parameter setting was run five times and the median performance was selected. Finally the optimal parameters that achieve the highest AUC for the brown target were selected and used for the other three target types.

### 5.1. Simulated data

As we discussed before, eFUMI combines all positive bags as one big positive bag and all negative bags as one big negative bag and learns target concept from the big positive bag that is different from the negative bag. So if the negative bags maintain incomplete knowledge of the background, e.g., some non-target concept appears only in the subset of positive bags, eFUMI will perform poorly. However, MI-HE which maintains bag structure will be able to distinguish the target.

Given this hypothesis, simulated data was generated from four spectra selected from the ASTER spectral library (Baldridge et al., 2009). Specifically, the Red Slate, Verde Antique, Phyllite and Pyroxenite spectra from the rock class with 211 bands and wavelengths ranging from  $0.4\text{ }\mu\text{m}$  to  $2.5\text{ }\mu\text{m}$  (as shown in Fig. 2 in solid lines) were used as endmembers to generate hyperspectral data. Red Slate was labeled as the target endmember.

Four sets of highly-mixed noisy data with varied mean target proportion value ( $\alpha_{t\_mean}$ ) were generated, a detailed generation process can be found in Jiao and Zare (2015). Specifically, this synthetic data has 15 positive and 5 negative bags with each bag having 500 points. If it is a positively labeled bag, there are 200 highly-mixed target points

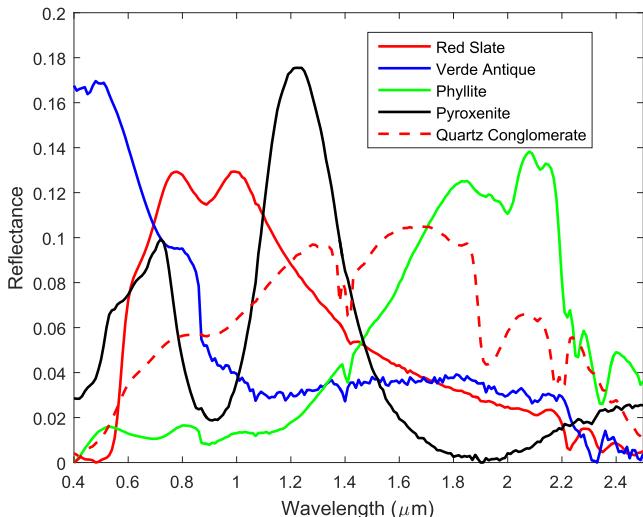


Fig. 2. Signatures from ASTER library used to generate simulated data.

Table 1

List of constituent endmembers for synthetic data with incomplete background knowledge.

Bag No.	Bag Label	Target Endmember	Background Endmember
1–5	+	Red Slate	Verde Antique, Phyllite, Pyroxenite
6–10	+	Red Slate	Phyllite, Pyroxenite
11–15	+	Red Slate	Pyroxenite
16–20	–	N/A	Phyllite, Pyroxenite

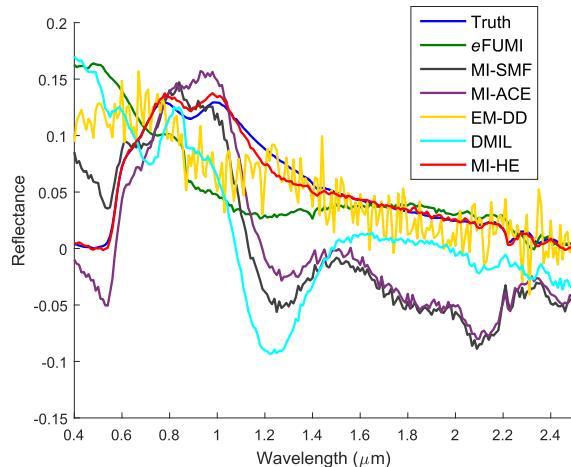
containing mean target (Red Slate) proportion from 0.1 to 0.7 respectively to vary the level of target presence from weak to high. Gaussian white noise was added so that signal-to-noise ratio of the data was set to 20 dB. To highlight the ability of MI-HE to leverage individual bag-level labels, we use different subsets of background endmembers to build synthetic data as shown in Table 1. Table 1 shows that the negatively labeled bags only contain 2 negative endmembers and there exists one confusing background endmember in the first 5 positive bags which is Verde Antique. It is expected that the proposed MI-HE will be able to learn the target signature correctly and eFUMI will confuse both Red Slate and Verde Antique as target signatures.

The parameter settings of MI-HE for this experiment are  $T = 1$ ,  $M = 9$ ,  $\rho = 0.8$ ,  $b = 5$ ,  $\beta = 5$  and  $\lambda = 1 \times 10^{-3}$ . MI-HE was compared to state-of-the-art MIL algorithms eFUMI (Zare and Jiao, 2014; Jiao and Zare, 2015), MI-SMF and MI-ACE (Zare et al., 2017), DMIL (Shrivastava et al., 2015, 2014), EM-DD (Zhang and Goldman, 2002) and mi-SVM (Andrews et al., 2002). The mi-SVM algorithm was added to these experiments to include a comparison MIL approach that does not rely on estimating a target signature.

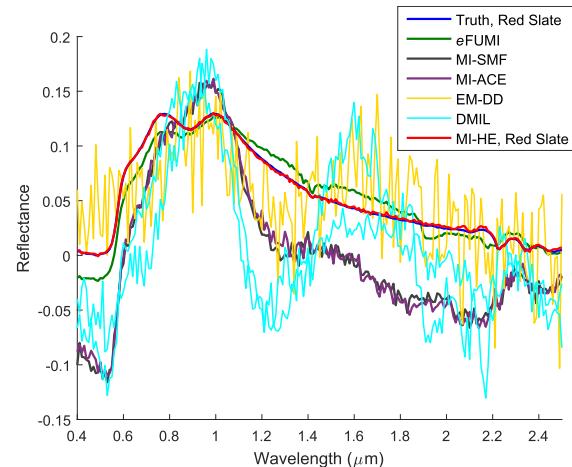
Fig. 3a shows the estimated target signature from data with 0.1 mean target proportion value. It clearly shows that the proposed MI-HE is able to correctly distinguish Red Slate as target concept from the incomplete background knowledge. Also, the other comparison algorithms can also estimate a target concept close to the groundtruth Red Slate spectrum. However, eFUMI is always confused with the other non-target endmember, Verde Antique, that exists in some positive bags but is excluded from the background bags.

For simulated detection analysis, estimated target concepts from the training data were then applied to the test data generated separately following the same generating procedure. The detection was performed using the HSD or ACE detection statistic. For MI-HE and eFUMI, both methods were applied since those two algorithms can come out a set of background concept from training simultaneously; for MI-SMF, both SMF and ACE were applied since MI-SMF's objective is maximizing the multiple instance spectral matched filter; for the rest multiple instance target concept learning algorithms, MI-ACE, EM-DD, DMIL, only ACE was applied. For the testing procedure of mi-SVM, a regular SVM testing process was performed using LIBSVM (Chang and Lin, 2011), and the decision values (signed distances to hyperplane) of test data determined from trained SVM model were taken as the confidence values. For the signature based detectors, the background data mean and covariance were estimated from the negative instances of the training data.

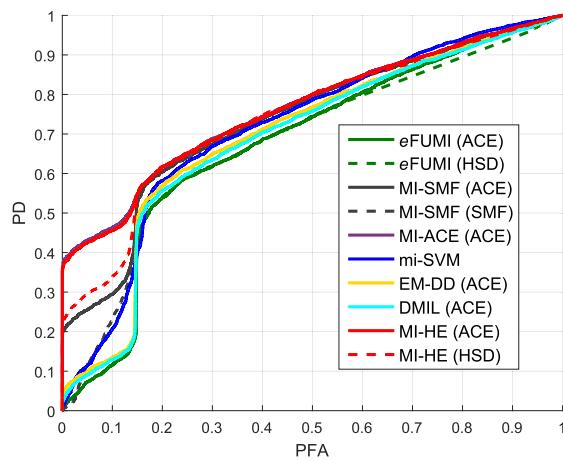
For quantitative evaluation, Fig. 3b shows the receiver operating characteristic (ROC) curves using estimated target signature, where it can be seen that the eFUMI is confused with the testing Verde Antique data at very low PFA (probability of false alarms) rate. Table 2 shows the area under the curve (AUC) of proposed MI-HE and comparison algorithms. The results reported are the median results over five runs of the algorithm on the same data. From Table 2, it can be seen that for MI-HE and MI-ACE, the best performance on detection was achieved using ACE detector, which is quite close to the performance of using the ground truth target signature (denoted as values with stars). The reason that MI-HE's detection using HSD detector is a little worse is that HSD relies on knowing the complete background concept to properly represent each non-target testing data, the missing non-target concept



(a) Estimated target signatures for Red Slate and comparison with ground truth



(a) Estimated target signatures for Red Slate and comparison with ground truth



(b) ROC curves cross validated on test data

**Fig. 3.** MI-HE and comparisons on synthetic data with incomplete background knowledge,  $\alpha_{t\_mean} = 0.1$ . MI-SMF and MI-ACE are not expected to recover the true signature.

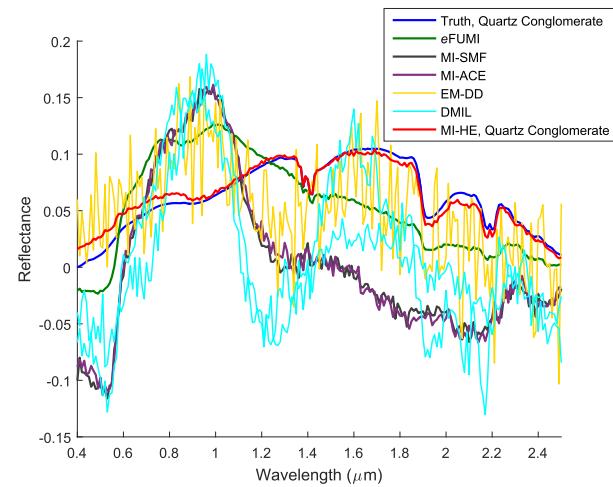
**Table 2**

Area under the ROC curves for MI-HE and comparison algorithms on Simulated Hyperspectral Data with Incomplete Background Knowledge. Best results shown in bold, second best results underlined, and ground truth shown with an asterisk.

Algorithm	$\alpha_{t\_mean}$			
	0.1	0.3	0.5	0.7
MI-HE (HSD)	0.743	<u>0.931</u>	0.975	0.995
MI-HE (ACE)	<u>0.763</u>	<u>0.952</u>	<u>0.992</u>	<u>0.999</u>
eFUMI (Jiao and Zare, 2015) (ACE)	0.675	0.845	0.978	0.998
eFUMI (Jiao and Zare, 2015) (HSD)	0.671	0.564	0.978	<u>0.998</u>
MI-SMF (Zare et al., 2017) (SMF)	0.719	0.923	0.972	0.993
MI-SMF (Zare et al., 2017) (ACE)	0.735	<u>0.952</u>	<u>0.992</u>	<u>0.999</u>
MI-ACE (Zare et al., 2017) (ACE)	<u>0.764</u>	<u>0.952</u>	<u>0.992</u>	<u>0.999</u>
mi-SVM (Andrews et al., 2002)	0.715	0.815	0.866	0.900
EM-DD (Zhang and Goldman, 2002) (ACE)	0.695	0.918	<u>0.983</u>	<u>0.998</u>
DMIL (Shrivastava et al., 2015, 2014) (ACE)	0.687	0.865	0.971	0.996
Ground Truth (ACE)	0.765*	0.953*	0.992*	0.999*

(Verde Antique) makes the non-target testing data containing Verde Antique maintain a relatively large reconstruction error, and thus large detection statistic.

To formulate a multiple concept learning problem, another rock

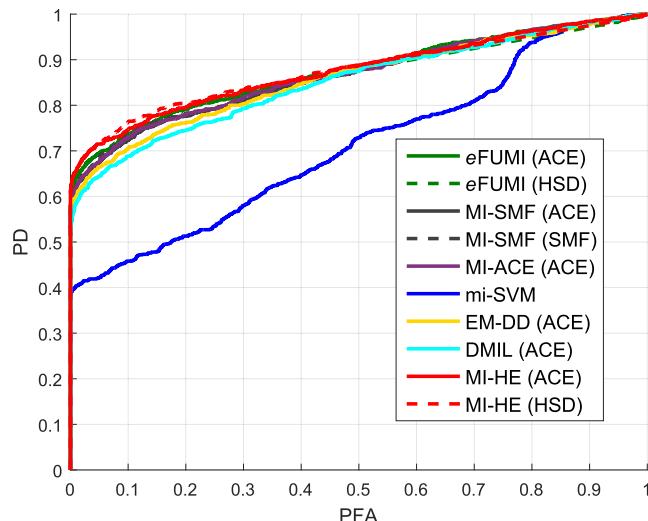


(a) Estimated target signatures for Quartz Conglomerate and comparison with ground truth

**Fig. 4.** Estimated target signatures by MI-HE and comparisons on synthetic data with multiple target concepts,  $\alpha_{t\_mean} = [0.1, 0.1]$ . Not all comparisons algorithms are expected to recover true target signatures.

endmember, Quartz Conglomerate, from ASTER spectral library was used as the second target concept (as shown in Fig. 2 in dashed line). Three sets of highly-mixed noisy data with varied mean target proportion value from [0.1, 0.1] to [0.3, 0.3] were generated. The synthetic data has 5 positive bags containing both target concept (Red Slate and Quartz Conglomerate) and non-target concept (Verde Antique, Phyllite, Pyroxenite); 5 negative bags containing only the background concept (Verde Antique, Phyllite, Pyroxenite). The mean target proportion value was set to [0.1, 0.1], [0.2, 0.2] and [0.3, 0.3], respectively to vary the level of target presence from weak to high. The other settings for simulated data are the same as above mentioned. It is expected that the proposed MI-HE is able to learn multiple target concept at one time.

The parameter settings of MI-HE for this experiment are  $T = 2$ ,  $M = 9$ ,  $\rho = 0.8$ ,  $b = 5$ ,  $\beta = 5$  and  $\lambda = 1 \times 10^{-3}$ . Fig. 4a and b show the estimated target concepts by proposed MI-HE and comparisons, where we can see that the proposed MI-HE is able to accurately estimate multiple target concepts simultaneously. Compared with MI-HE, although DMIL is also a multiple concept learning algorithm, target concept as estimated by DMIL is noisy, and not a representative prototype of the target class. The remaining comparison algorithms are single target concept learning which are always confused by the multiple target concept problem.

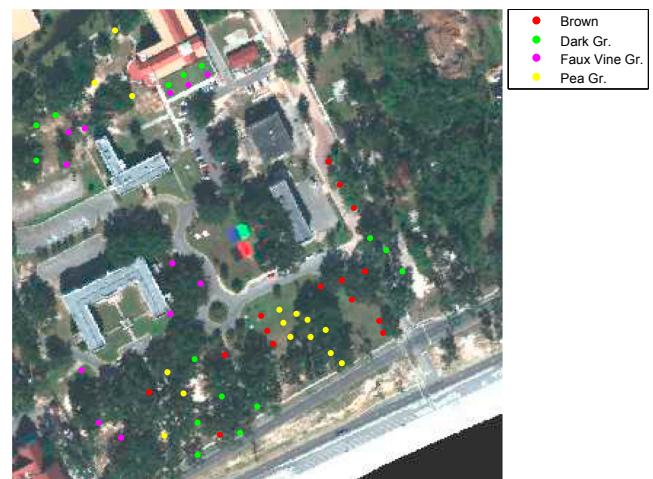


**Fig. 5.** ROCs of MI-HE and comparisons on synthetic data with multiple target concepts,  $\alpha_{t\_mean} = [0.1, 0.1]$ .

Fig. 5 shows the ROCs using estimated target signature, and Table 3 shows the AUCs of proposed MI-HE and comparison algorithms. The results reported are the median results over five runs of the algorithm on the same data. Since ACE is a single signature based detector, given multiple target concepts, the maximum detection statistic across all estimated target concept was picked for each testing data. From Table 3, it can be seen that the proposed MI-HE outperforms all the comparison single concept learning algorithms as well as the multiple MI concept learning algorithm DMIL. The reason MI-HE even outperforms detection performance using ground truth signatures is that there also exists target points which are mixtures of two target concepts. The hybrid detector can directly model the mixture of target concepts outperforming the application of only one concept at a time.

## 5.2. MUUFL Gulfport hyperspectral data

For experiments on real hyperspectral target detection data, the MUUFL Gulfport hyperspectral data set collected over the University of Southern Mississippi-Gulfpark Campus was used. This data set contains  $325 \times 337$  pixels with 72 spectral bands corresponding to wavelengths from 367.7 nm to 1043.4 nm at a 9.5–9.6 nm spectral sampling interval with spatial resolution 1 pixel/ $m^2$  (Gader and Zare, 2013). The first four and last four bands were removed due to sensor noise. Two sets of this data (Gulfport Campus Flight 1 and Gulfport Campus Flight 3) were



**Fig. 6.** MUUFL Gulfport data set RGB image and the 57 target locations.

selected as cross-validated training and testing data for these two data sets have the same altitude and spatial resolution. Throughout the scene, there are 64 man-made targets in which 57 were considered in this experiment which are cloth panels of four different colors: Brown (15 examples), Dark Green (15 examples), Faux Vineyard Green (FVGr) (12 examples) and Pea Green (15 examples). The spatial location of the targets are shown as scattered points over an RGB image of the scene in Fig. 6. Some of the targets are in the open ground and some are occluded by the live oak trees. Moreover, the targets also vary in size, for each target type, there are targets that are  $0.25 m^2$ ,  $1 m^2$  and  $9 m^2$  in area, respectively, resulting a very challenging, highly mixed sub-pixel target detection problem.

### 5.2.1. MUUFL Gulfport hyperspectral data, individual target type detection

For this part of the experiments, each individual target type was treated as a target class, respectively. For example, when “Brown” is selected as target class, a  $5 \times 5$  rectangular region corresponding to each of the 15 ground truth locations denoted by GPS was grouped into a positive bag to account for the drift coming from GPS. This size was chosen based on the accuracy of the GPS device used to record the ground truth locations. The remaining area that does not contain a brown target was grouped into a big negative bag. This constructs the detection problem for “Brown” target. Similarly, there are 15, 12, 15 positive labeled bags for Dark Green, Faux Vineyard Green and Pea Green, respectively. The parameter settings of MI-HE for this experiment are  $T = 1$ ,  $M = 9$ ,  $\rho = 0.3$ ,  $b = 5$ ,  $\beta = 1$  and  $\lambda = 5 \times 10^{-3}$ .

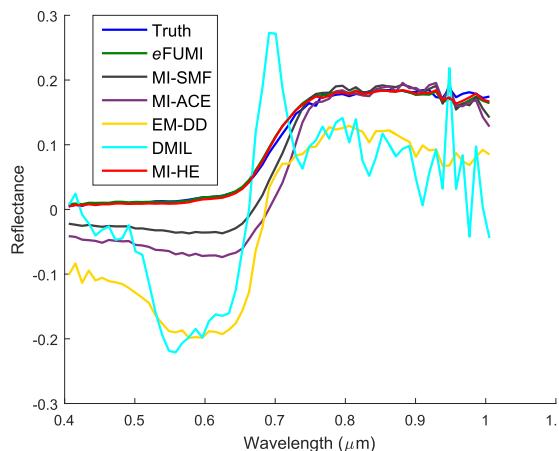
MI-HE and comparison algorithms were evaluated on this data using the Normalized Area Under the receiver operating characteristic Curve (NAUC) in which the area was normalized out to a false alarm rate (FAR) of  $1 \times 10^{-3}$  false alarms/ $m^2$  (Glenn et al., 2013). During detection on the test data, the background mean and covariance were estimated from the negative instances of the training data. The results reported are the median results over five runs of the algorithm on the same data.

Fig. 7a shows the estimated target concept by proposed MI-HE and comparisons for Brown target type training on flight 3. We can see that the proposed MI-HE is able to recover the target concept quite close to groundtruth spectra manually selected from the scene. Fig. 7b shows the detection ROCs given target spectra estimated on flight 3 and cross validated on flight 1. Table 4 shows the NAUCs for MI-HE and comparison algorithms cross validated on all four types of target, where it can be seen the proposed MI-HE generally outperforms the comparisons for most of the target types and achieves close to the performance of using ground truth target signatures. Since MI-HE is a discriminative target concept learning framework that aims to distinguish one target instance from each positively labeled bag, MI-HE had a lower performance for the pea green target because of the relatively large occlusion

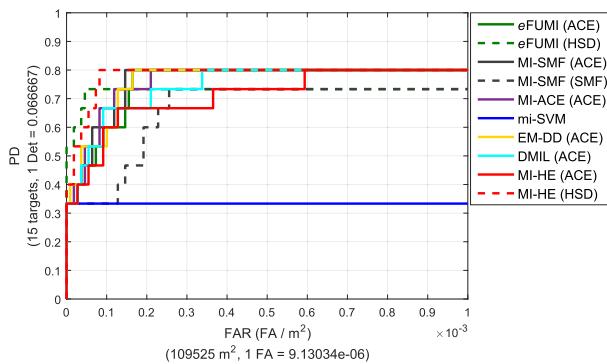
**Table 3**

Area under the ROC curves for MI-HE and comparison algorithms on Simulated Hyperspectral Data with Multiple Target Concepts. Best results shown in bold, second best results underlined, and ground truth shown with an asterisk.

Algorithm	$\alpha_{t\_mean}$		
	[0.1, 01]	[0.2, 0.2]	[0.3, 0.3]
MI-HE (HSD)	<b>0.875</b>	<b>0.982</b>	<b>0.998</b>
MI-HE (ACE)	<b>0.875</b>	<b>0.982</b>	<b>0.998</b>
eFUMI (Jiao and Zare, 2015) (ACE)	<u>0.872</u>	<u>0.980</u>	<b>0.998</b>
eFUMI (Jiao and Zare, 2015) (HSD)	0.865	0.976	<u>0.997</u>
MI-SMF (Zare et al., 2017) (SMF)	0.866	0.977	<u>0.997</u>
MI-SMF (Zare et al., 2017) (ACE)	0.865	0.976	<u>0.997</u>
MI-ACE (Zare et al., 2017) (ACE)	0.866	0.976	<u>0.997</u>
mi-SVM (Andrews et al., 2002)	0.711	0.890	0.970
EM-DD (Zhang and Goldman, 2002) (ACE)	0.858	0.979	<b>0.998</b>
DMIL (Shrivastava et al., 2015, 2014) (ACE)	0.850	0.971	0.994
Ground Truth (ACE)	0.869*	0.979*	0.997*



(a) Estimated target signatures from flight 3 for Brown and comparison with ground truth



(b) ROC curves cross validated on flight 1

**Fig. 7.** MI-HE and comparisons on Gulfport Data Brown, training flight 3 testing flight 1.

of those targets causing difficulty in distinguishing pea green signature from each of the positive bag.

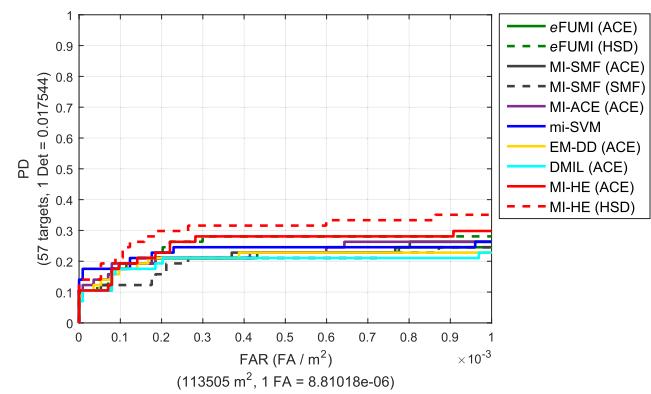
#### 5.2.2. MUUFL Gulfport hyperspectral data, all four target types detection

For training and detection for the four target types together, the positive bags were generated by grouping each of the  $5 \times 5$  regions denoted by the groundtruth that it contains any of the four types of target. Thus, for each flight there are 57 target points and 57 positive bags were generated. The remaining area that does not contain any target was grouped into a big negative bag. The parameter settings of MI-HE for this experiment are  $T = 9$ ,  $M = 11$ ,  $\rho = 0.3$ ,  $b = 5$ ,  $\beta = 1$  and

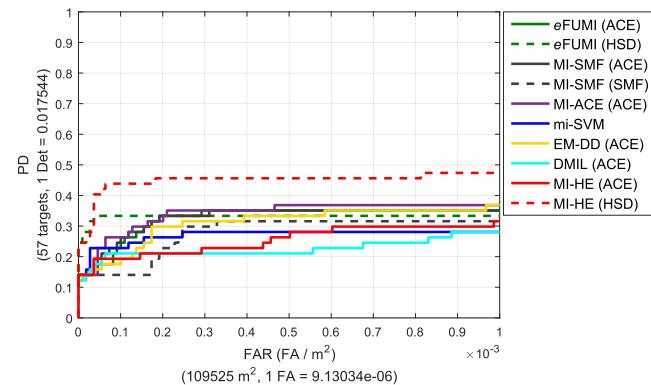
**Table 4**

Area under the ROC curves for MI-HE and comparison algorithms on Gulfport Data with Individual Target Type. Best results shown in bold, second best results underlined, and ground truth shown with an asterisk.

Alg.	Train on Flight 1; Test on Flight 3				Train on Flight 3; Test on Flight 1			
	Brown	Dark Gr.	Faux Vine Gr.	Pea Gr.	Brown	Dark Gr.	Faux Vine Gr.	Pea Gr.
MI-HE (HSD)	<b>0.499</b>	<b>0.453</b>	<b>0.655</b>	0.267	<b>0.781</b>	<b>0.532</b>	<b>0.655</b>	0.350
MI-HE (ACE)	0.433	0.379	0.104	0.267	0.710	0.360	0.111	0.266
eFUMI (Jiao and Zare, 2015) (ACE)	0.423	0.377	<u>0.654</u>	0.267	0.754	0.491	0.605	<u>0.393</u>
eFUMI (Jiao and Zare, 2015) (HSD)	<b>0.444</b>	<u>0.436</u>	0.653	0.267	0.727	<u>0.509</u>	0.500	0.333
MI-SMF (Zare et al., 2017) (SMF)	0.419	0.354	0.533	0.266	0.657	0.405	<u>0.650</u>	0.384
MI-SMF (Zare et al., 2017) (ACE)	0.448	0.382	0.579	<u>0.316</u>	<b>0.760</b>	0.501	0.613	0.388
MI-ACE (Zare et al., 2017) (ACE)	<u>0.474</u>	0.390	0.485	<u>0.333</u>	<u>0.760</u>	0.483	0.593	0.380
mi-svm (Andrews et al., 2002)	0.206	0.195	0.412	0.265	0.333	0.319	0.245	0.274
EM-DD (Zhang and Goldman, 2002) (ACE)	0.411	0.381	0.486	0.279	<b>0.760</b>	0.503	0.541	<b>0.416</b>
DMIL (Shrivastava et al., 2015, 2014) (ACE)	0.419	0.383	0.191	0.009	0.743	0.310	0.081	0.083
Ground Truth (ACE)	0.528*	0.429*	0.656*	0.267*	0.778*	0.521*	0.663*	0.399*



(a) Train on flight 1, detect on flight 3



(b) Train on flight 3, detect on flight 1

**Fig. 8.** ROCs of MI-HE and comparisons on Gulfport Data all types detection.

$$\lambda = 5 \times 10^{-3}$$

**Fig. 8a** and **b** shows the detection ROCs given target spectra estimated one flight and cross validated on another flight, which show that the detection statistic by proposed MI-HE using HSD is significantly better than the comparison algorithms. **Table 5** summarizes the NAUCs as a quantitative comparison.

#### 5.3. Tree species classification from NEON data

In this section, we applied MI-HE to a multiple instance tree species classification problem using a subset of National Ecological

**Table 5**

Area under the ROC curves for MI-HE and comparison algorithms on Gulfport Data with All Four Target Types. Best results shown in bold, second best results underlined, and ground truth shown with an asterisk.

Alg.	Tr. Fl. 1; Te. Fl. 3	Tr. Fl. 3; Te. Fl. 1	Alg.	Tr. Fl. 1; Te. Fl. 3	Tr. Fl. 3; Te. Fl. 1
MI-HE (HSD)	<b>0.304</b>	<b>0.449</b>	MI-SMF (Zare et al., 2017) (ACE)	0.219	0.327
MI-HE (ACE)	<u>0.257</u>	0.254	MI-SMF (Zare et al., 2017) (SMF)	0.198	0.277
eFUMI (Jiao and Zare, 2015) (ACE)	0.214	<u>0.325</u>	mi-SVM (Andrews et al., 2002)	0.235	0.269
eFUMI (Jiao and Zare, 2015) (HSD)	0.256	0.331	EM-DD (Zhang and Goldman, 2002) (ACE)	0.211	0.310
MI-ACE (Zare et al., 2017) (ACE)	0.226	0.340	DMIL (Shrivastava et al., 2015, 2014) (ACE)	0.198	0.225
Ground Truth (ACE)	0.330*	0.490*			

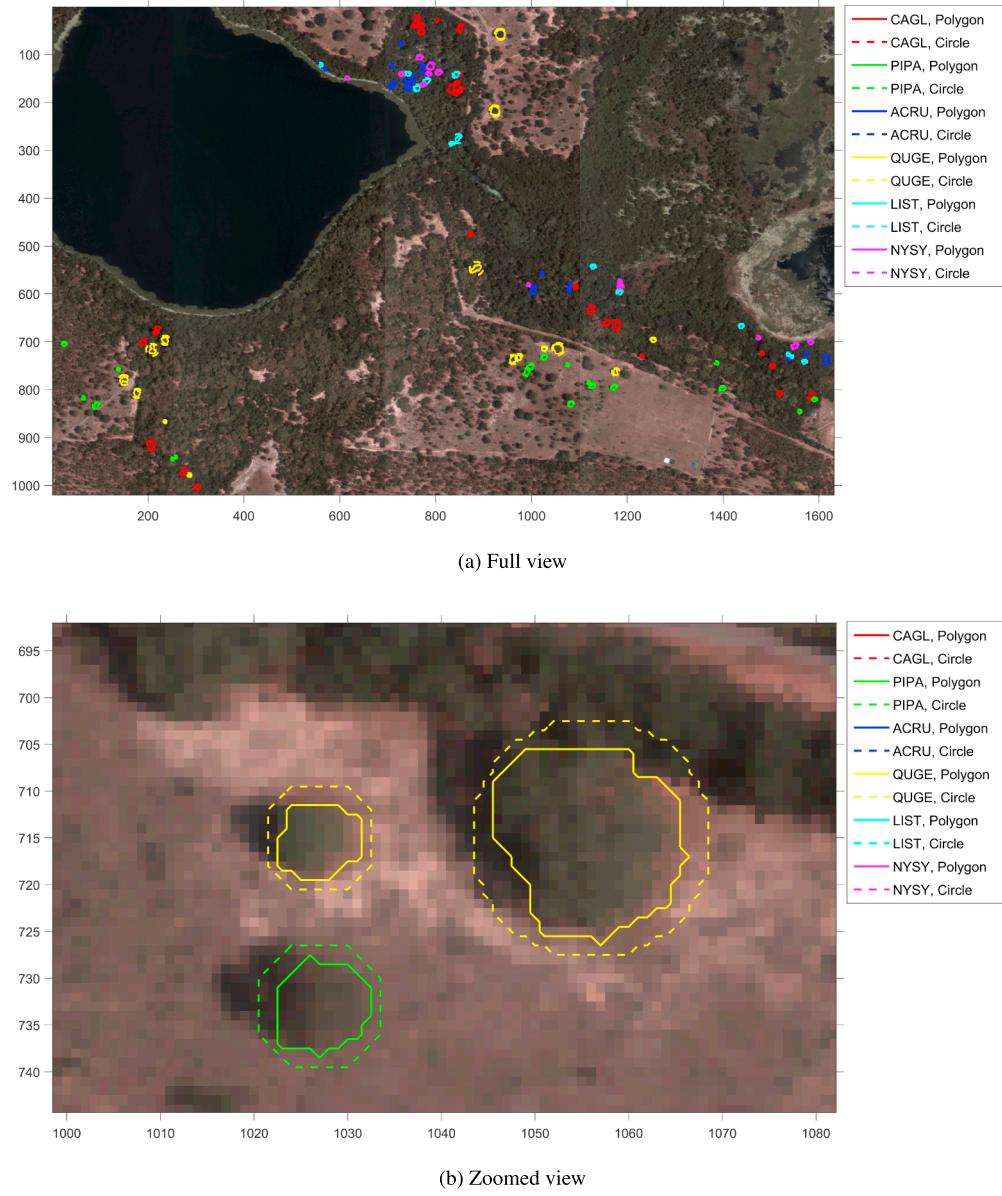
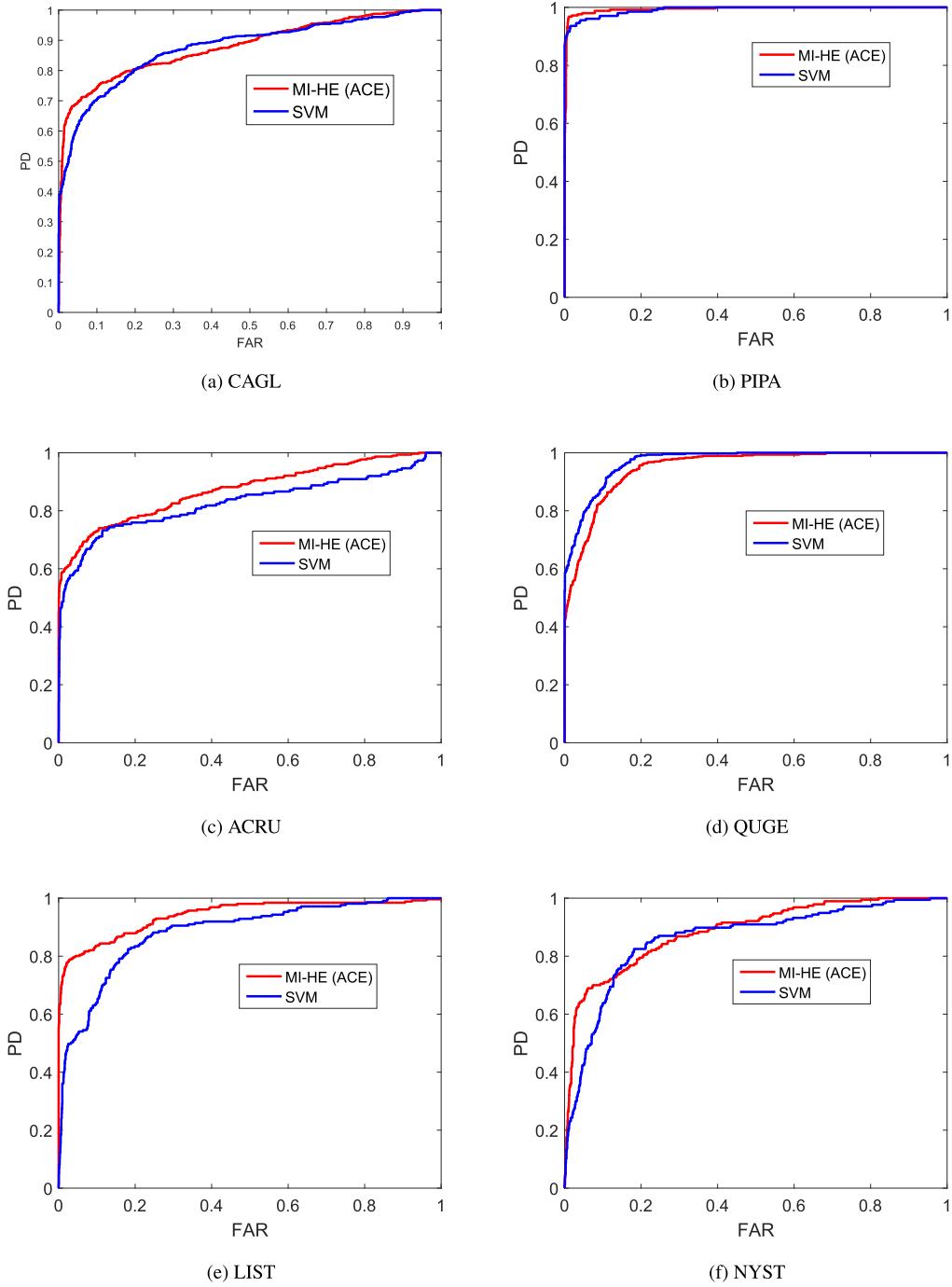


Fig. 9. RGB image of NEON OSBS with tree polygons.

Observatory Network (NEON) hyperspectral data ([National Ecological Observatory Network, 2016](#)) collected at the Ordway-Swisher Biological Station (OSBS) in north-central Florida, United States. This data contains  $1020 \times 1631$  pixels with 428 bands corresponding to wavelengths from 380 nm to 2510 nm at a 5 nm spectral sampling interval. The spatial resolution and collection altitude are 1 pixel/m<sup>2</sup> and 1000 m, respectively.

Fig. 9a shows the full ground view (RGB) of the hyperspectral data

where there are six types of trees named CAGL (*Carya glabra*), PIPA (*Pinus palustris*), ACRU (*Acer rubrum*), QUGE (*Quercus geminata*), LIST (*Liquidambar styraciflua*) and NYSY (*Nyssa sylvatica*), denoted by polygons with different colors. The number of labeled trees for each type is: 26, 18, 16, 15, 13 and 13, respectively. Fig. 9b shows a zoomed in view of Fig. 9a, where the solid line denotes the polygon boundaries of each tree's crown. The polygon boundaries were mapped in the field on the image loaded on a GPS-connected tablet. These polygon contours create



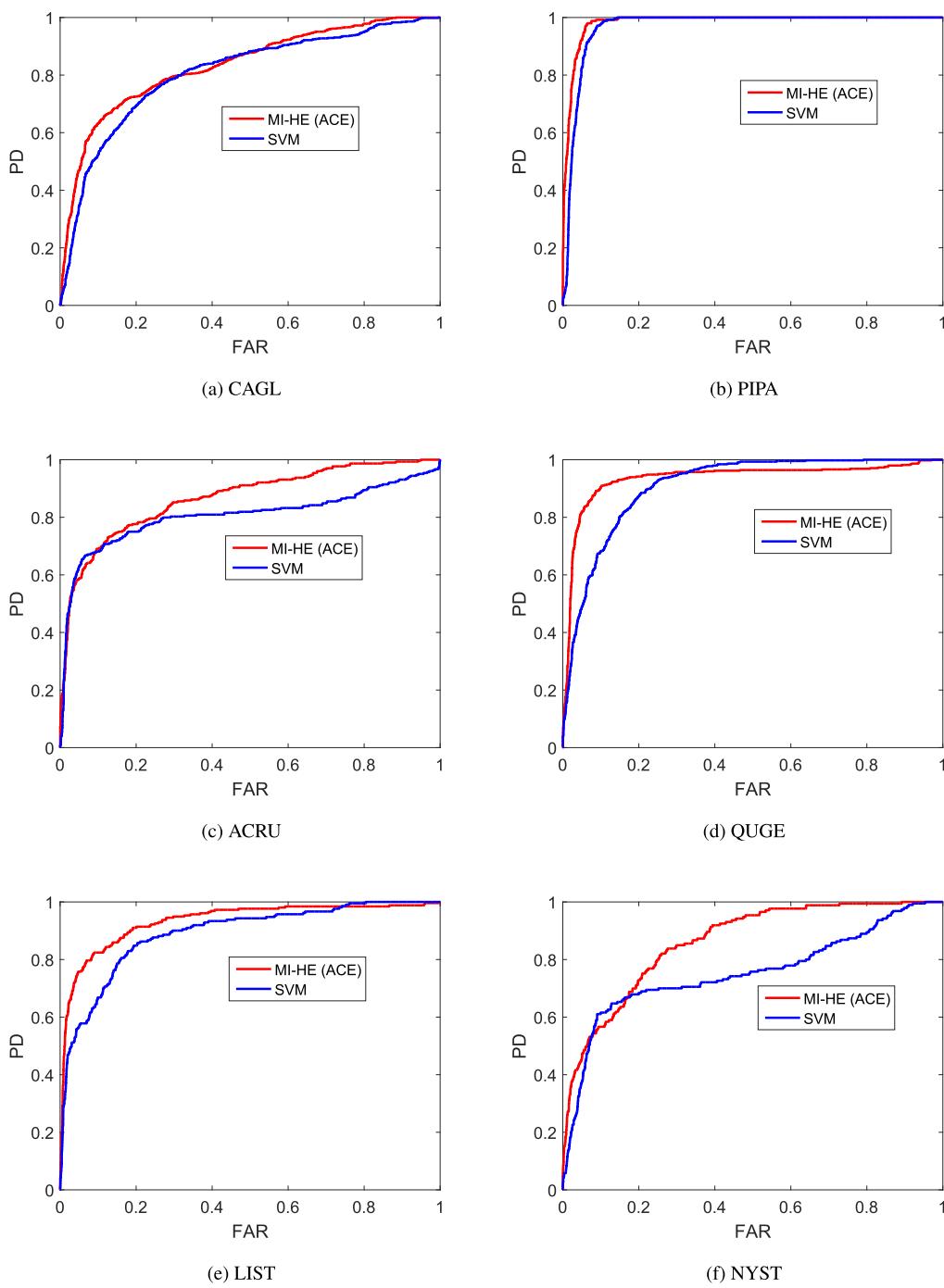
**Fig. 10.** ROC curves of MI-HE and SVM on polygon data.

an accurately labeled training data set for tree species classification. Several existing methods (Graves et al., 2016; Nia et al., 2015) train a SVM from crown polygons for tree species classification. However, obtaining accurate tree canopy polygons is time consuming, requiring a lot of effort in the field. Furthermore, the canopy labels are inherently inaccurate, e.g., the accuracy of GPS can drift several meters; the boundary of a tree crown can be ambiguous.

So here we model this task as a MIL problem and more general, inaccurate labels were generated by drawing a circle including each polygon shown as dashed line in Fig. 9b. The goal is to show the proposed MI-HE is able to do tree species classification well given the inaccurate circle labels. One could imagine these “circle labels” being generated using the GPS coordinate of the center of a tree and an approximate tree radius, which would be much easier to collect than the

precise polygon labels. The experiment was conducted by training on 70% of the canopies, randomly selected, and testing on the remaining 30%. The testing step was conducted by scoring the testing data point by point, where the testing data were per-pixel labeled according to the polygon data. Hierarchical dimensionality reduction was applied to reduce the dimensionality of the data to 124 (Zare et al., 2018). Both the circle data and polygon data were used as training and compared with SVM. The parameter settings of MI-HE for this experiment are  $T = 1$ ,  $M = 8$ ,  $\rho = 0.5$ ,  $b = 5$ ,  $\beta = 1$  and  $\lambda = 1 \times 10^{-3}$ . The experiments were repeated for five times and the median performance (AUC) was shown.

Figs. 10 and 11 show the ROC curves of MI-HE and SVM that train and test on polygon data and circle data, respectively. Table 6 shows the detailed AUCs for each run. For training and testing on the polygon



**Fig. 11.** ROC curves of MI-HE and SVM on circle data.

**Table 6**

Tree Species Classification Results (area under the ROC curves) by MI-HE and SVM, Best results shown in bold.

Species	Polygon Data		Circle Data	
	MI-HE	SVM	MI-HE	SVM
CAGL	<b>0.880</b>	0.876	<b>0.833</b>	0.807
PIPA	<b>0.994</b>	0.992	<b>0.984</b>	0.969
ACRU	<b>0.873</b>	0.829	<b>0.867</b>	0.806
QUGE	0.950	<b>0.970</b>	<b>0.936</b>	0.910
LIST	<b>0.942</b>	0.884	<b>0.933</b>	0.890
NYSY	<b>0.887</b>	0.862	<b>0.861</b>	0.755
Average	<b>0.921</b>	0.902	<b>0.902</b>	0.856

data, SVM outperforms MI-HE only on tree type QUGE and has close to but lower overall performance than MI-HE. However, for training and testing on the circle data, MI-HE outperforms SVM on each type of the classification. Furthermore, compared with the results from the polygon data, although MI-HE and SVM both provide decreased performance on the circle data, SVM suffers more from the label uncertainty given the circle data.

## 6. Analysis of MI-HE parameter settings on simulated data

In order to provide deeper insights into the sensitivity of MI-HE performance relative to variations in input parameters, we tested MI-HE on simulated hyperspectral data across a range of parameter values. Specifically, the varying parameters and ranges examined are shown in

**Table 7**  
Testing ranges for analysis of MI-HE parameter settings.

Type	Range
$M$	1, 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21
$\beta$	0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100
$\lambda$	$1 \times 10^{-4}, 2 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 2 \times 10^{-3}, 5 \times 10^{-3}$ , 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1
$b$	-10, -5, -2, -1, $1 \times 10^{-10}$ , 1, 2, 5, 10, 20, 50, 100

**Table 8**  
List of constituent endmembers for synthetic data.

Bag No.	Bag Label	Constituent Endmembers
1–5	+	Red Slate, Verde Antique, Phyllite, Pyroxenite
6–10	-	Verde Antique Phyllite, Pyroxenite

**Table 7.** Red Slate was again selected as target endmember and the other three, Verde Antique, Phyllite and Pyroxenite were used as non-target endmembers as shown in Fig. 2. Table 8 shows the bags labeling and constituent endmembers. The synthetic data has  $K^+ = 5$  positive and  $K^- = 5$  negative bags with each bag containing 100 points. If it is a positively labeled bag, there are 50 highly-mixed target points containing mean target (Red Slate) proportion  $\alpha_{t\_mean} = 0.1$  and are mixed with at least one randomly picked background endmember. Gaussian white noise was added so that the signal-to-noise ratio of the data was set to 20 dB. The reference parameters for MI-HE were set to  $T = 1$ ,  $M = 7$ ,  $\rho = 0.8$ ,  $b = 5$ ,  $\beta = 5$  and  $\lambda = 1 \times 10^{-3}$ .

Fig. 12 shows the detection performance (mean AUCs and variance

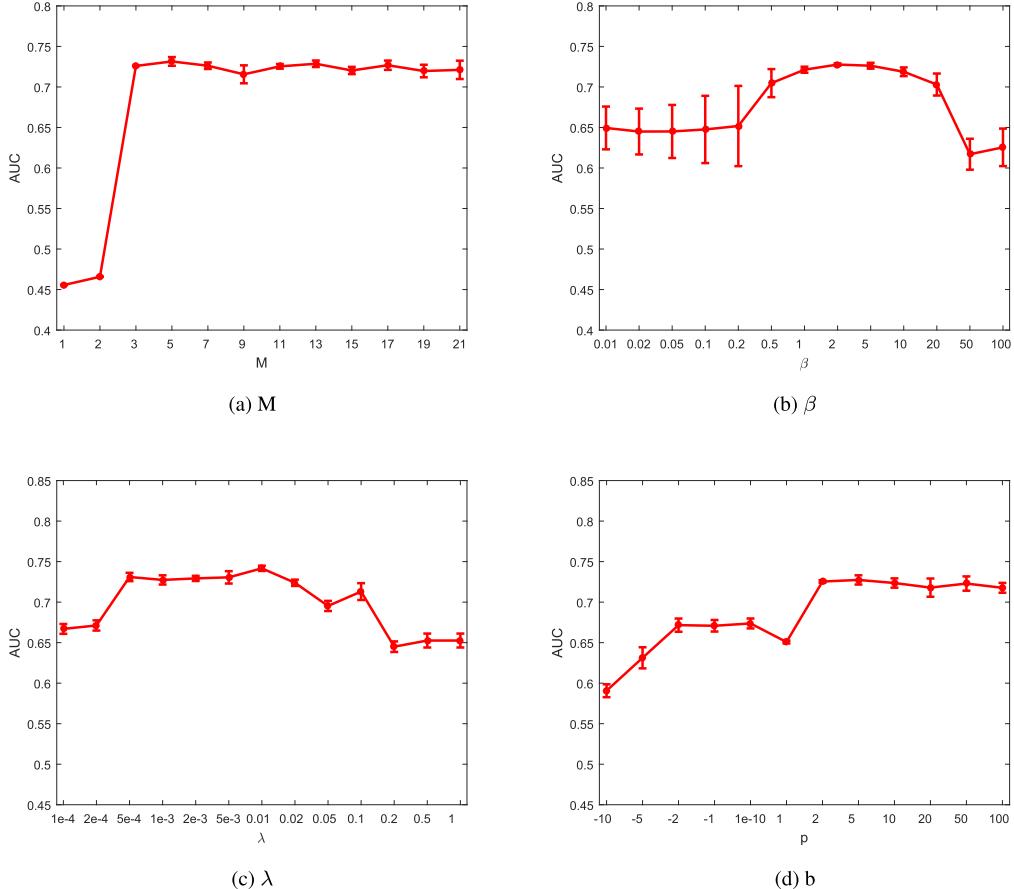
**Table 9**  
Testing ranges for MI-HE 4D parameter perturbation testing.

Type	Range
$M$	3, 5, 7, 9
$\beta$	1, 2, 5, 10
$\lambda$	$1 \times 10^{-3}, 2 \times 10^{-3}, 5 \times 10^{-3}$ , 0.01
$b$	5, 10, 20, 50

over five runs) of MI-HE for the sensitivity analysis on the simulated data described above. Several interesting inferences can be drawn from Fig. 12 regarding how MI-HE responds to its parameter settings. For the setting of  $M$ , it can be seen from Fig. 12a that MI-HE performs consistently when the number of background concepts is set to greater than or equal to 3, which is the true number of background concepts of the simulated data. Since MI-HE adopts a linear mixing model instead of convex mixing for the data mixture, the conclusion is MI-HE is not sensitive to the setting of  $M$  given  $M$  is not set smaller than the necessary constituent background concepts of input data.

For the setting of  $\beta$ , Fig. 12b shows MI-HE performs well with  $\beta$  in the range [1, 10] on this data. Since  $\beta$  is a scaling factor for the hybrid detector,  $Pr(\mathbf{x}_{ij}|\mathbf{D}, \mathbf{B}_i^+) = \exp\left(-\beta \frac{\|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}^+\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^T\mathbf{p}_{ij}\|^2}\right)$ , the scaling effect controls the gradient of the objective function (10). Since the range  $\beta \in [1, 10]$  provides a moderate gradient for the exponential function, so it can be concluded that MI-HE requires  $\beta$  to be set to a suitable range, e.g., [1, 10].

The setting of sparsity level,  $\lambda$ , results in the step length of soft shrinkage. This value should be related with the magnitude of the input data. For the simulated data tested here, the proportion values were generated from the Dirichlet distribution within range [0, 1]. As shown



**Fig. 12.** Detection statistics (AUCs) of MI-HE plots with different parameter settings.

by Fig. 12c, the appropriate range of  $\lambda$  for this simulated data is  $[5 \times 10^{-4}, 0.02]$  which is reasonable. However, in general, prior knowledge is needed for setting specific  $\lambda$  for a real dataset. Currently we set  $\lambda$  to be approximately the 1/1000 of the  $l_2$ -norm mean of the training data.

The parameter  $b$  controls the sharpness of the generalized mean model to realize the operations from minimum ( $b \rightarrow -\infty$ ), average ( $b \in (0, 1]$ ) to maximum ( $b \rightarrow +\infty$ ). Since the proposed model aims to emphasize the most true positive instance from each positive bag and assumes the “soft maximum” operation for this generalized mean model, it is expected that the model will work well with  $b$  greater than 1. The setting of this  $b$  value was discussed in (Wu et al., 2014; Shrivastava et al., 2015), where  $b$  was set to 1.5 (Wu et al., 2014) and 10 (Shrivastava et al., 2015) and observed to work well. Fig. 12d verifies this hypothesis showing that the algorithm works well for  $b$  great than 1. For all experiments shown in this paper, the parameter  $b$  was set to 5.

In order to further evaluate the robustness of MI-HE to its parameters, a 4D parameters perturbation testing was conducted, in which the parameters were varied across all the possible combinations of the four parameters. The range is listed in Table 9 which is a subset of Table 7. The best and worst performance (median AUC over five runs) from this 4D loop are 0.742 and 0.695, respectively. The difference between the best and worst performance in percentage is 6.24%, which further validates MI-HE’s stability on a large range of parameters.

Although there are several parameters for MI-HE, these parameters come from the models assumed to underlie MI-HE. The sensitivity analysis of MI-HE parameters on a very noisy, highly mixed synthetic data provides a solid verification for the model stability. It can be concluded that the proposed MI-HE is generally robust to the

perturbations of its parameters and there is a general range of model stability for the space of the MI-HE parameters. Moreover, the above analysis provides a intuitive understanding and heuristic approach for setting the parameters of MI-HE to stable ranges.

## 7. Conclusion

In this work the MI-HE target concept learning framework for MIL problems is proposed and investigated. MI-HE is able to learn multiple discriminative target concepts from ambiguously labeled data. After learning target concepts, target detection can be conducted by applying the estimated target concepts to any signature based detector. Comprehensive experiments show that the proposed MI-HE is effective in learning discriminative target concept and achieves superior performance over comparison algorithms in several scenarios. Although this work was inspired by sub-pixel hyperspectral target detection, the MI-HE algorithm is a general MIL concept learning algorithm that could be applied to problems with mixed and ambiguously labeled training data.

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1350078 - CAREER: Supervised Learning for Incomplete and Uncertain Data, and the National Natural Science Foundation of China under Grant No. 61806157. The NEON crown polygon data was collected as part of National Science Foundation Dimensions of Biodiversity Grant No. 1442280.

## Appendix A

Similar to the Dictionary Learning using Singular Value Decomposition (K-SVD) approach (Aharon et al., 2006), the optimization of target and background concepts is performed by taking gradient descent with respect to one atom at a time and holding the rest fixed. Denote  $f_{GM}$  as the generalized mean part of Eq. (10):

$$\begin{aligned} f_{GM} &= - \sum_{i=1}^{K^+} \frac{1}{b} \ln \left( \frac{1}{N_i} \sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i)^b \right) \\ &= - \sum_{i=1}^{K^+} \frac{1}{b} \ln \frac{1}{N_i} - \sum_{i=1}^{K^+} \frac{1}{b} \ln \left( \sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i)^b \right) \end{aligned} \quad (\text{A.1})$$

Removing the constant part in  $f_{GM}$  and taking the partial derivative with respect to  $\mathbf{d}$  results in (A.2):

$$\frac{\partial f_{GM}}{\partial \mathbf{d}} = - \sum_{i=1}^{K^+} \frac{1}{\sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i)^b} \cdot \left( \sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i)^{b-1} \cdot \frac{\partial \Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i)}{\partial \mathbf{d}} \right), \quad (\text{A.2})$$

where  $\mathbf{d}$  is a symbolic notation for any atom in  $\mathbf{D}$ .

Then take the partial derivative on the fidelity (second) term of the objective function Eq. (10) with respect to the background concept,  $\mathbf{d}_k$ :

$$\begin{aligned} &\frac{\partial -\rho \sum_{i=K^++1}^K \sum_{j=1}^{N_i} \ln Pr(l_{ij} = -1 | \mathbf{B}_i)}{\partial \mathbf{d}_k} \\ &= -\rho \sum_{i=K^++1}^K \sum_{j=1}^{N_i} \frac{\| \mathbf{x}_{ij} - \mathbf{D}^{-1} \mathbf{p}_{ij} \|_2^2}{\sum_{k=1}^{N_i} p_{ijk}} \\ &= \rho \sum_{i=K^++1}^K \sum_{j=1}^{N_i} -2p_{ijk} (\mathbf{x}_{ij} - \mathbf{D}^{-1} \mathbf{p}_{ij}) \\ &= \rho \sum_{i=K^++1}^K \sum_{j=1}^{N_i} -2p_{ijk} \mathbf{q}_{ij}, \end{aligned} \quad (\text{A.3})$$

where  $p_{ijk}$  is the  $k$ th element in  $\mathbf{p}_{ij}$  corresponding to  $\mathbf{d}_k$ .

The partial derivative of the cross incoherence (third) term corresponding to  $\mathbf{d}_t$  is:

$$\frac{\partial Q(\mathcal{X}, \mathbf{D}^+, \mathcal{A})}{\partial \mathbf{d}_t} = \alpha \sum_{i=K^++1}^K \sum_{j=1}^{N_i} (\mathbf{D}^+ \mathbf{a}_{ij}^+)^T \mathbf{x}_{ij} \cdot \mathbf{a}_{ij}^+ \mathbf{x}_{ij}$$
(A.4)

The partial derivatives of the negative objective function Eq. (10) with respect to  $\mathbf{d}_t$  and  $\mathbf{d}_k$  are shown in (A.5) and (A.6).

$$\begin{aligned} \frac{\partial J_3}{\partial \mathbf{d}_t} = & - \sum_{i=1}^{K^+} \frac{1}{\sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)^b} \\ & \cdot \left( \sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)^{b-1} \cdot \frac{\partial \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)}{\partial \mathbf{d}_t} \right) \\ & + \alpha \sum_{i=K^++1}^K \sum_{j=1}^{N_i} (\mathbf{D}^+ \mathbf{a}_{ij}^+)^T \mathbf{x}_{ij} \cdot \mathbf{a}_{ij}^+ \mathbf{x}_{ij} \end{aligned}$$
(A.5)

$$\begin{aligned} \frac{\partial J_3}{\partial \mathbf{d}_k} = & - \sum_{i=1}^{K^+} \frac{1}{\sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)^b} \\ & \cdot \left( \sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)^{b-1} \cdot \frac{\partial \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)}{\partial \mathbf{d}_k} \right) \\ & + \rho \sum_{i=K^++1}^K \sum_{j=1}^{N_i} -2p_{ijk} \mathbf{q}_{ij} \end{aligned}$$
(A.6)

The next step is taking the partial derivative of the hybrid detector in (6) with respect to  $\mathbf{d}_t$  and  $\mathbf{d}_k$  shown as Eqs. (A.7) and (A.8) respectively:

$$\begin{aligned} \frac{\partial \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)}{\partial \mathbf{d}_t} = & \exp \left( -\beta \frac{\|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^{-}\mathbf{p}_{ij}\|^2} \right) \frac{\partial \left( \frac{-\beta \|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^{-}\mathbf{p}_{ij}\|^2} \right)}{\partial \mathbf{d}_t} \\ = & \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i) \frac{2\beta \mathbf{a}_{ij}^+ (\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij})}{\|\mathbf{x}_{ij} - \mathbf{D}^{-}\mathbf{p}_{ij}\|^2} \\ = & \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i) \frac{2\beta \mathbf{a}_{ij}^+ \mathbf{r}_{ij}}{\|\mathbf{q}_{ij}\|^2} \end{aligned}$$
(A.7)

$$\begin{aligned} \frac{\partial \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)}{\partial \mathbf{d}_k} = & \exp \left( -\beta \frac{\|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^{-}\mathbf{p}_{ij}\|^2} \right) \frac{\partial \left( \frac{-\beta \|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^{-}\mathbf{p}_{ij}\|^2} \right)}{\partial \mathbf{d}_k} \\ = & \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i) \frac{2\beta \mathbf{a}_{ijk}^+ \mathbf{r}_{ij} \|\mathbf{q}_{ij}\|^2 - 2\beta p_{ijk} \|\mathbf{r}_{ij}\|^2 \mathbf{q}_{ij}}{\|\mathbf{q}_{ij}\|^4} \end{aligned}$$
(A.8)

Substituting the gradient of hybrid detector with respect to  $\mathbf{d}_t$  and  $\mathbf{d}_k$  in Eqs. (A.7) and (A.8) to Eqs. (A.5) and (A.6), respectively, we can get the resultant gradient of the objective function (10) over  $\mathbf{d}_t$  and  $\mathbf{d}_k$ :

$$\begin{aligned} \Delta \mathbf{d}_t = & - \sum_{i=1}^{K^+} \frac{1}{\sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)^b} \\ & \left( \sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)^b \cdot \frac{2\beta \mathbf{a}_{ij}^+ \mathbf{r}_{ij}}{\|\mathbf{q}_{ij}\|^2} \right) \\ & + \alpha \sum_{i=K^++1}^K \sum_{j=1}^{N_i} (\mathbf{D}^+ \mathbf{a}_{ij}^+)^T \mathbf{x}_{ij} \cdot \mathbf{a}_{ij}^+ \mathbf{x}_{ij} \end{aligned}$$
(A.9)

$$\begin{aligned} \Delta \mathbf{d}_k = & - \sum_{i=1}^{K^+} \frac{1}{\sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)^b} \\ & \left( \sum_{j=1}^{N_i} \Lambda(\mathbf{x}_{ij}, \mathbf{D} | \mathbf{B}_i)^b \cdot 2\beta \frac{a_{ijk}^+ \mathbf{r}_{ij} \|\mathbf{q}_{ij}\|^2 - p_{ijk} \|\mathbf{r}_{ij}\|^2 \mathbf{q}_{ij}}{\|\mathbf{q}_{ij}\|^4} \right) \\ & - \rho \sum_{i=K^++1}^K \sum_{j=1}^{N_i} 2p_{ijk} \mathbf{q}_{ij} \end{aligned}$$
(A.10)

The optimization of sparse representation can be viewed as a  $l_1$  regularized least squares problem, also known as lasso problem (Tibshirani, 1996; Chen et al., 2001), denoted as  $\mathcal{L}$ . The lasso problem is shown in Eq. (5), where given concept (or dictionary) set  $\mathbf{D}$  and preset regularization constant  $\lambda$ ,  $\mathbf{a}^*$  is the optimal sparse representation of the input data  $\mathbf{x}$ . Here we adopt the iterative shrinkage-thresholding algorithm (ISTA) (Figueiredo and Nowak, 2003; Daubechies et al., 2004) for solving for the sparse codes  $\mathbf{a}$ .

The gradient of (5) with respect to  $\mathbf{a}$  without considering the  $l_1$  penalty term is:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a}} = -\mathbf{D}^T(\mathbf{x} - \mathbf{Da}). \quad (\text{A.11})$$

Then  $\mathbf{a}$  at  $q^{th}$  iteration can be updated using gradient descent shown in (A.12):

$$\mathbf{a}^q = \mathbf{a}^{q-1} - \delta \frac{\partial \mathcal{L}}{\partial \mathbf{a}}, \quad (\text{A.12})$$

followed by a soft-thresholding step:

$$\mathbf{a}^* = S_\lambda(\mathbf{a}^q), \quad (\text{A.13})$$

where  $S_\lambda: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is the *soft-thresholding* operator defined by

$$S_\lambda(\mathbf{a}[k]) = \text{sign}(\mathbf{a}[k]) \max(|\mathbf{a}[k]| - \lambda, 0), k = 1, \dots, n \quad (\text{A.14})$$

Following a similar proof to that in (Facchinei and Pang, 2007), when the step length  $\delta$  satisfies (A.15), the update of  $\mathbf{a}$  using a gradient descent method with step length  $\eta$  monotonically decreases the value of the objective function, where  $Eig_{\max}(\mathbf{D}^T \mathbf{D})$  denotes the maximum eigenvalue of  $\mathbf{D}^T \mathbf{D}$ . For simplicity,  $\delta$  was set to  $(0.99/(Eig_{\max}(\mathbf{D}^T \mathbf{D}))$  for all input data:

$$\delta \in \left(0, \frac{1}{Eig_{\max}(\mathbf{D}^T \mathbf{D})}\right) \quad (\text{A.15})$$

Finally the resultant update equation for the sparse representation of instance  $\mathbf{x}$  given concept set  $\mathbf{D}$  is:

$$\mathbf{a}^* = S_\lambda\left(\mathbf{a}^q + \frac{1}{Eig_{\max}(\mathbf{D}^T \mathbf{D})}(\mathbf{D}^T(\mathbf{x} - \mathbf{Da}^q))\right) \quad (\text{A.16})$$

## References

- Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD: an algorithm for designing over-complete dictionaries for sparse representation. *IEEE Trans. on Signal Process.* 54 (11), 4311–4322.
- Andrews, S., Tsochantaridis, I., Hofmann, T., 2002. Support vector machines for multiple-instance learning. In: Advances Neural Inf. Process. Syst. (NIPS), pp. 561–568.
- Babenko, B., Dollar, P., Tu, Z., Belongie, S., 2008. Simultaneous learning and alignment: Multi-instance and multi-pose learning. In: Workshop on Faces in ‘Real-Life’ Images: Detection, Alignment, and Recognition.
- Bach, F., Jenatton, R., Mairal, J., Obozinski, G., 2012. Optimization with sparsity-inducing penalties. *Found. Trends Mach. Learn.* 4 (1), 1–106.
- Baldridge, A., Hook, S., Grove, C., Rivera, G., 2009. The ASTER spectral library version 2.0. *Remote Sens. Environ.* 113 (4), 711–715.
- Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., et al., 2012. and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 5 (2), 354–379.
- Broadwater, J., Chellappa, R., 2007. Hybrid detectors for subpixel targets. *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (11), 1891–1903.
- Broadwater, J., Meth, R., Chellappa, R., 2004. A hybrid algorithm for subpixel detection in hyperspectral imagery. In: Proc. IEEE Intl. Geosci. Remote Sens. Symp. (IGARSS), vol. 3, pp. 1601–1604.
- Bullen, P., Mitrinović, D., Vasić, P., 1988. Means and their inequalities. *Math. Appl. Chang*, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2 (3), 1–27.
- Chen, S.S., Donoho, D.L., Saunders, M.A., 2001. Atomic decomposition by basis pursuit. *SIAM Rev.* 43 (1), 129–159.
- Chen, Y., Bi, J., Wang, J.Z., 2006. MILES: Multiple-instance learning via embedded instance selection. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (12), 1931–1947.
- Daubecies, I., Defrise, M., De Mol, C., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Applied Math.* 57, 1413–1457.
- Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T., 1997. Solving the multiple instance problem with axis-parallel rectangles. *Artif. Intell.* 89 (1), 31–71.
- Eismann, M.T., Stocker, A.D., Nasrabadi, N.M., 2009. Automated hyperspectral cueing for civilian search and rescue. *Proc. IEEE* 97 (6), 1031–1055.
- Facchinei, F., Pang, J.-S., 2007. Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer Science & Business Media.
- Figueiredo, M.A., Nowak, R.D., 2003. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.* 12 (8), 906–916.
- Gader, P., Zare, A., et al. (2013). MUUFL gulfport hyperspectral and lidar airborne data set. Tech. rep., University of Florida, Gainesville, FL, REP-2013-570.
- Glenn, T., Zare, A., Gader, P., Dranishnikov, D., 2013. Bullwinkle: scoring code for sub-pixel targets (version 1.0) [software]. <<http://engineers.missouri.edu/zarea/code/>>.
- Graves, S.J., Asner, G.P., Martin, R.E., Anderson, C.B., Colgan, M.S., Kalantari, L., Bohlman, S.A., 2016. Tree species abundance predictions in a tropical agricultural landscape with a supervised classification model and imbalanced data. *Remote Sens.* 8 (2), 161.
- Jiao, C., Zare, A., 2015. Functions of multiple instances for learning target signatures. *IEEE Trans. Geosci. Remote Sens.* 53 (8), 4670–4686.
- Jiao, C., Zare, A., 2016. Multiple instance dictionary learning using functions of multiple instances. In: Int. Conf. Pattern Recognition (ICPR), pp. 2688–2693.
- Jiao, C., Zare, A., 2017. Multiple instance hybrid estimator for learning target signatures. In: Proc. IEEE Intl. Geosci. Remote Sens. Symp. (IGARSS), pp. 1–4.
- Jiao, C., Zare, A., 2018. GatorSense/MIHE: Initial release (version 0.1), Zenodo. <https://doi.org/10.5281/zenodo.1320109>.
- Keshava, N., Mustard, J.F., 2002. Spectral unmixing. *IEEE Signal Process. Mag.* 19 (1), 44–57.
- Kraus, O.Z., Ba, J.L., Frey, B.J., 2016. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics* 32 (12), i52–i59.
- Kraut, S., Scharf, L., 1999. The CFAR adaptive subspace detector is a scale-invariant GLRT. *IEEE Trans. Signal Process.* 47 (9), 2538–2541.
- Kraut, S., Scharf, L., McWhorter, L., 2001. Adaptive subspace detectors. *IEEE Trans. Signal Process.* 49 (1), 1–16.
- Landgrebe, D., 2002. Hyperspectral image data analysis. *IEEE Signal Process. Mag.* 19 (1), 17–28.
- Lara, M., Lleó, L., Diezma-Iglesias, B., Roger, J.-M., Ruiz-Altisent, M., 2013. Monitoring spinach shelf-life with hyperspectral image through packaging films. *J. Food Eng.* 119 (2), 353–361.
- Mahajan, G., Sahoo, R., Pandey, R., Gupta, V., Kumar, D., 2014. Using hyperspectral remote sensing techniques to monitor nitrogen, phosphorus, sulphur and potassium in wheat (*triticum aestivum* l.). *Precis. Agric.* 15 (5), 499–522.
- Mairal, J., Bach, F., Ponce, J., 2014. Sparse modeling for image and vision processing. *Found. Trends Comput. Graph. Vision* 8 (2–3), 85–283.
- Mallat S., 2008. A wavelet tour of signal processing. Third ed.: The Sparse Way, Academic Press.
- Manolakis, D., Shaw, G., 2002. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Process. Mag.* 19 (1), 29–43.
- Maron, O., Lozano-Perez, T., 1998. A framework for multiple-instance learning. In: Advances Neural Inf. Process. Syst. (NIPS), vol. 10, pp. 570–576.
- Maron, O., Ratan, A.L., 1998. Multiple-instance learning for natural scene classification. In: Int. Conf. Mach. Learning (ICML), vol. 98, pp. 341–349.
- Nascimento, J.M., Dias, J.M., 2005. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* 43 (4), 898–910.
- Nasrabadi, N.M., 2008. Regularized spectral matched filter for target recognition in hyperspectral imagery. *IEEE Signal Process. Lett.* 15, 317–320. <https://doi.org/10.1109/LSP.2008.917805>.
- Nasrabadi, N.M., 2014. Hyperspectral target detection: an overview of current and future challenges. *IEEE Signal Process. Mag.* 31 (1), 34–44.
- National Ecological Observatory Network, 2016. Data accessed on Jan., 2016. Available on-line <<http://data.neonscience.org/>> from Battelle, Boulder, CO, USA.
- Nia, M.S., Wang, D.Z., Bohlman, S.A., Gader, P., Graves, S.J., Petrovic, M., 2015. Impact of atmospheric correction and image filtering on hyperspectral classification of tree species using support vector machine. *J. Appl. Remote Sens.* 9 (1), 095990.
- Pardo, A., Real, E., Krishnaswamy, V., et al., 2017. Directional kernel density estimation for classification of breast tissue spectra. *IEEE Trans. Med. Imag.* 36 (1), 64–73.
- Pike, R., Lu, G., Wang, D., et al., 2016. A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging. *IEEE Trans. Biomed. Eng.* 63 (3), 653–663.
- Quellec, G., Lamard, M., Cozic, M., Coatrieux, G., Cazuguel, G., 2016. Multiple-instance learning for anomaly detection in digital mammography. *IEEE Trans. Med. Imag.* 35 (7), 1604–1614.
- Ramirez, I., Sprechmann, P., Sapiro, G., 2010. Classification and clustering via dictionary

- learning with structured incoherence and shared features. In: IEEE Conf. Comput. Vision Pattern Recognition (CVPR), pp. 3501–3508.
- Shrivastava, A., Pillai, J.K., Patel, V.M., Chellappa, R., 2014. Dictionary-based multiple instance learning. In: IEEE Int. Conf. Image Process., pp. 160–164.
- Shrivastava, A., Patel, V.M., Pillai, J.K., Chellappa, R., 2015. Generalized dictionaries for multiple instance learning. *Int. J. of Comput. Vision* 114 (2), 288–305.
- Theiler, J., Foy, B.R., 2006. Effect of signal contamination in matched-filter detection of the signal on a cluttered background. *IEEE Geosci. Remote Sens. Lett* 3 (1), 98–102. <https://doi.org/10.1109/LGRS.2005.857619>.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. Series B (Methodol.)* 267–288.
- Wang, Z., Lan, L., Vucetic, S., 2012. Mixture model for multiple instance regression and applications in remote sensing. *IEEE Trans. Geosci. Remote Sens.* 50 (6), 2226–2237.
- Wohlhart, P., Köstinger, M., Roth, P.M., Bischof, H., 2011. Multiple instance boosting for face recognition in videos. In: Joint Pattern Recognition Symposium. Springer, pp. 132–141.
- Wu, J., Zhao, Y., Zhu, J.-Y., Luo, S., Tu, Z., 2014. Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pp. 256–263.
- Xu, Y., Zhu, J.-Y., Chang, E., Tu, Z., 2012. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 964–971.
- Yang, M., Zhang, L., Feng, X., Zhang, D., 2011. Fisher discrimination dictionary learning for sparse representation. In: Int. Conf. Comput. Vision. IEEE, pp. 543–550.
- Yang, M., Zhang, L., Feng, X., Zhang, D., 2014. Sparse representation based fisher discrimination dictionary learning for image classification. *Int. J. Comput. Vision* 109 (3), 209–232.
- Yuksel, S.E., Bolton, J., Gader, P., 2015. Multiple-instance hidden markov models with applications to landmine detection. *IEEE Trans. Geosci. Remote Sens.* 53 (12), 6766–6775.
- Zare, A., Jiao, C., 2014. Extended functions of multiple instances for target characterization. In: IEEE Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS), pp. 1–4.
- Zare, A., Bolton, J., Gader, P., Schatten, M., 2008. Vegetation mapping for landmine detection using long-wave hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* 46 (1), 172–178. <https://doi.org/10.1109/TGRS.2007.906438>.
- Zare, A., Jiao, C., Glenn, T. Discriminative multiple instance hyperspectral target characterization. *IEEE Trans. Pattern Anal. Mach. Intell.*
- Zare, A., Glenn, T., Gader, P., 2018. Gatorsense/hsi\_toolkit (version v0.1), Zenodo. <https://doi.org/10.5281/zenodo.1186417>.
- Zhang, Q., Goldman, S.A., 2002. EM-DD: an improved multiple-instance learning technique. In: Advances Neural Inf. Process. Syst. (NIPS), vol. 2, pp. 1073–1080.
- Zhu, J., Rosset, S., Hastie, T., Tibshirani, R., 2004. 1-norm support vector machines. In: Advances Neural Inf. Process. Syst. (NIPS), vol. 16, pp. 49–56.