

Multiple Instance Learning for Multiple Diverse Hyperspectral Target Characterizations

Ping Zhong^{ID}, Senior Member, IEEE, Zhiqiang Gong^{ID}, and Jiaxin Shan

Abstract—A practical hyperspectral target characterization task estimates a target signature from imprecisely labeled training data. The imprecisions arise from the characteristics of the real-world tasks. First, accurate pixel-level labels on training data are often unavailable. Second, the subpixel targets and occluded targets cause the training samples to contain mixed data and multiple target types. To address these imprecisions, this paper proposes a new hyperspectral target characterization method to produce diverse multiple hyperspectral target signatures under a multiple instance learning (MIL) framework. The proposed method uses only bag-level training samples and labels, which solves the problems arising from the mixed data and lack of pixel-level labels. Moreover, by formulating a multiple characterization MIL and including a diversity-promoting term, the proposed method can learn a set of diverse target signatures, which solves the problems arising from multiple target types in training samples. The experiments on hyperspectral target detections using the learned multiple target signatures over synthetic and real-world data show the effectiveness of the proposed method.

Index Terms—Diversity, human in loop, hyperspectral data, multiple instance learning (MIL), multiple target characterizations.

I. INTRODUCTION

HYPERSPECTRAL images, which provide a dense spectral sampling at each pixel, have proven useful in the literature of remote sensing, computer vision, and machine learning [9], [33], [45], [46]. As an important hyperspectral image analysis task, hyperspectral target characterization estimates a signature to effectively represent targets from given training samples. Such characterization is critical for many hyperspectral data applications, particularly for object detection [28], [33], [41] and classification [1], [11], [12], [34], [40], [43], where the learned target spectral signature is typically used as a template for the detectors or classifiers. However, the target characterization is a challenging hyperspectral image analysis task. There are three common

Manuscript received May 29, 2018; revised November 29, 2018 and January 7, 2019; accepted February 12, 2019. This work was supported in part by the National Science Foundation of China under Grant 61671456 and Grant 61271439, in part by the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201243, and in part by the Program for New Century Excellent Talents in University under Grant NECT-13-0164. (*Corresponding author: Ping Zhong*.)

The authors are with the National Key Laboratory of Science and Technology on ATR, College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: zhongping@nudt.edu.cn; gongzhiqiang13@nudt.edu.cn; sjx5104@yahoo.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNNLS.2019.2900465

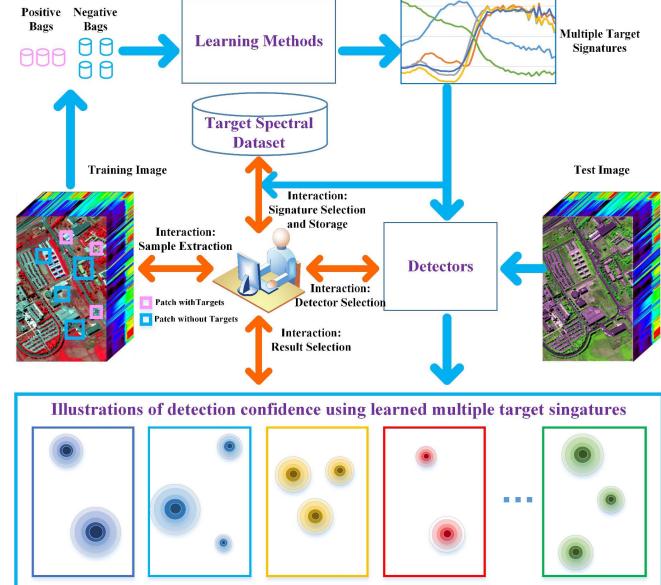


Fig. 1. Interactive detection system, where the human is added into the loop.

approaches used to obtain the target signatures, including laboratory testing, collection by hand-held spectrometers, and extraction from real-world hyperspectral images. However, these common methods could be ineffective because they do not account for atmospheric or environmental conditions in the laboratory and handheld spectrometer measured signatures, as well as the difficulty in identifying the pure target pixel when extracting signal from real-world hyperspectral images [41].

To address these challenges, a feasible practical workflow that can obtain the effective target signatures is designed, as shown in Fig. 1. The workflow is actually performed in many real-world image analysis systems, which can be briefly described as follows. The user selects the training image patches and assigns the patches labels denoting whether a patch contains the given target type. Then, input the training patches into a learning algorithm to produce target signature candidates. At the testing procedure, the test hyperspectral images and learned signature candidates are sent to an object detector and output multiple detection results. Finally, the signatures that produce the ideal result evaluated by the user are put into storage for subsequent applications. The designed workflow can solve the challenges

from the previously mentioned common target characterization approaches. First, the target characterization is performed over real-world hyperspectral images and thus can effectively consider effects from environmental and atmospheric conditions. Second, the flowchart needs only patch-level training samples and their labels and thus avoids identifying pixel-level targets.

However, there are also several issues needed to be addressed when working in the designed flowchart. First, only patch-level training samples and their labels are available. Second, the subpixel targets and occluded targets cause the training samples contain mixed data and multiple target types. Third, since there are multiple target types, the learning method is required to have the ability to learn multiple target signatures, which should be diverse enough to guarantee their discriminatory powers. To address all these challenges, this paper proposes a new hyperspectral target characterization method to produce multiple diverse target signatures under multiple instance learning (MIL) framework.

A. Contributions

Our contributions are threefold. First, we develop a multiple characterization MIL method that uses only bag-level training samples to learn signatures from mixed data and thus could solve the problems arising from mixed data and the lack of pixel-level labels. Second, by including a diversity-promoting cost in the objective function of multiple characterizations MIL, the proposed method learns a set of diverse target signatures and thus could solve the problems arising from multiple target types in training samples. Third, extensive experiments over both synthetic and real-world data demonstrate the method's practicability and its superiority over the state-of-the-art methods.

B. Related Work

In fact, the learning with imprecise and noisy labels is a very popular topic in the literature [2], [10], [13], [18], [19], [31]. In this paper, the proposed method learns target signatures from imprecise labels, which is closely related to multiple instance concept learning. MIL was mainly developed for classifications [4]–[7], [14], [29], [39] and regressions [21], [37]. A few MIL methods were developed to learn concepts, including diverse density (DD) [32], expectation-maximization DD (EM-DD) [42], MIL for spectral matched filter (MIL-SMF) and adaptive cosine estimator (MIL-ACE) [41], diversified dictionaries for MIL [35], multiple instance dictionary learning using functions or extended functions of multiple instances (eFUMI) [23]–[25], and multiple instance hybrid estimator (MI-HE) [22], [26].

DD, EM-DD, MIL-SMF, and MIL-ACE tend to estimate a target concept that lies close to at least one instance in each positive bag and maximizes the distance from all instances in negative bags. These methods differ in the distance metric employed. DD and EM-DD use the Euclidean distance [32], [42], whereas MIL-SMF and MIL-ACE use cosine similarity [41]. These methods can learn only a single target signature and are not appropriate for the problems discussed in this paper.

An eFUMI estimates a set of concepts, instead of learning a single target concept close to the conjunction of positive bags and far from each negative instance. However, the set of concepts has only one concept for the target, whereas other concepts are for the background. In theory, we can set one of the target types in the given training samples as target and other target types as background, which could allow us to use eFUMI to learn multiple target concepts. However, eFUMI typically places importance on modeling the target type. Other target types set as background could be neglected and cannot be effectively represented.

MI-HE proposed in [22] and [26] aims to learn multiple target concepts through introducing a data mixing model and optimizing response of a hybrid subpixel detector within a MIL framework. The proposed method can simultaneously learn multiple target and background signatures. The strength of the proposed method partially derives from the unmixing in MIL framework, which makes the proposed method very effective when a target is mixed with nontargets or other considering target types.

The method proposed in this paper also aims to learn multiple target signatures but is different from MI-HE in the following aspects. First, our method does not explicitly model target and background signals and thus focuses mainly on learning target signatures. Second, our method is derived totally from a kind of real-world application task, where the selected positive bags could have multiple target types that do not appear in negative bags. This makes the proposed multiple characterization MIL to treat multiple targets completely in MIL framework. Finally, the real-world application task, which motivates our method, allows human in the loop and thus has efficiency requirements on learning procedure. Our method uses a simple gradient-based method to optimize objective function and thus is relatively efficient.

II. MULTIPLE CONCEPTS IN MIL TARGET CHARACTERIZATIONS

This paper aims to develop an easy operational and relative low time-consuming workflow to obtain the target characterization for hyperspectral object detection. A feasible workflow is designed as follows: the hyperspectral image patches with/without the considering objects are manually selected, respectively, and then the selected image patches and their labels denoting with/without objects are used as input for a learning algorithm. The workflow is operationally easy and efficient since it needs only locating image patches and corresponding patch-level labels. However, a patch-level label presents only the information on whether a patch contains the considering object type and cannot indicate whether a pixel in the patch is a target or background instance. This situation makes the learning of target characterization weakly supervised and challenging. It can be solved by the MIL method [41], where an image patch with or without objects is named as a positive or negative bag, respectively, and each pixel sample in bags is called instance.

The usual MIL target characterization methods could learn an appropriate target signature if only one target type is

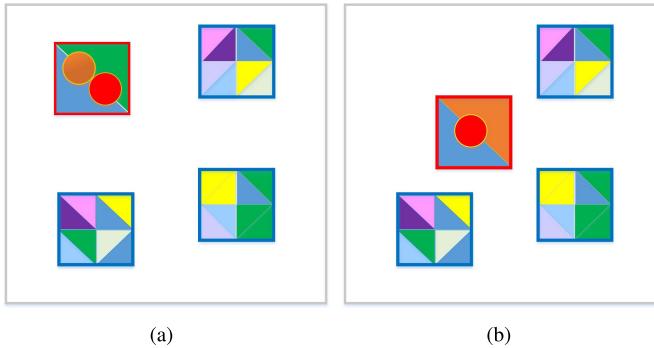


Fig. 2. Multiple concepts in the MIL hyperspectral target characterization method. (a) and (b) Two situations where the positive bag contains two kinds of targets.

presented in the given training samples [41]. However, under the feasible workflow depicted in Fig. 1, it is usual to select the nonideal positive and negative bags, which contain multiple target types, and thus multiple concepts needed to be considered in MIL. Fig. 2 shows two usual situations where the selected training samples contain multiple concepts. The selected positive and negative bags are image patches labeled with red and blue boxes, respectively. Other blank sites do not participate in the MIL. In each selected image patches, the pixels with different colors denote different types of targets and background land-covers. For simplicity, we use the circles and triangles to denote the considered targets and background land-covers. Only one positive bag and three negative bags are presented in the figure.

Fig. 2(a) demonstrates that the selected positive bag could contain two kinds of targets (labeled as orange and red color) and two kinds of background land-covers (labeled as blue and green color). Comparing colors in both positive and negative bags, we can find that the negative bags do not contain any orange and red color instances, and thus both target types in positive bag should be treated as the candidates to produce final target characterizations. Fig. 2(b) further shows that although only one kind of target is presented in the selected positive bag, the land-cover labeled as orange color should be treated as a kind of candidate target since all the negative bags do not contain any orange color land-cover. Therefore, the usual MIL target characterization methods, which are effective to learn only one concept, are not suitable for situations presented in Fig. 2. It is necessary to develop a new method to learn multiple concepts simultaneously. Moreover, multiple concepts should be diverse to cover different kinds of targets. All of these motivate us to develop a new diverse multiple target characterization method.

III. DIVERSE MULTIPLE TARGET CHARACTERIZATIONS

A. Multiple Characterization MIL

The problem can be formulated as follows. Let $B = \{B_1, B_2, \dots, B_M\}$ be a group of M selected hyperspectral image patches. Each patch B_m is named as a bag and the spectral vector $\mathbf{x} = [x_1, x_2, \dots, x_d] \in B_m$ denotes a instance in bag B_m , where d is the number of spectral bands. The

binary bag-level labels $L = \{L_1, L_2, \dots, L_M\}$ where $L_m \in \{0, 1\}$ are assigned to each bag to demonstrate whether the bag contains the target. A positive bag (B_m with $L_m = 1$) that contains at least one instance composed of the target is denoted as B_m^+ , whereas a negative bag (B_m with $L_m = 0$) that contains entirely nontarget instances is B_m^- . Assume that there are K target types in the given training bags. The framework to learn the K characterizations is generally written as

$$\mathbf{S}^* = \arg \max_{\mathbf{S}} \prod_m P(\mathbf{S}|B_m^+) \prod_m P(\mathbf{S}|B_m^-) \quad (1)$$

where $\mathbf{S} = \langle \mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(K)} \rangle$ is the K -tuple of target signatures, and $P(\mathbf{S}|B_m^+)$ and $P(\mathbf{S}|B_m^-)$ denote the probabilities of the signatures given the positive and negative bags, respectively.

The probabilities $P(\mathbf{S}|B_m^+)$ and $P(\mathbf{S}|B_m^-)$ can be defined to capture the characteristics of training data. Because this paper investigates the target characterization for detection, the probabilities are modeled to simultaneously capture characteristics of both training data and subsequent target detection. Derived from (1), the new detector specific MIL for multiple target characterizations can be equivalently written as

$$\mathbf{S}^* = \arg \max_{\mathbf{S}} \{C_1(\mathbf{S}) + C_2(\mathbf{S})\}, \quad (2)$$

$$C_1(\mathbf{S}) = \frac{1}{N^+} \sum_{m:L_m=1} \Omega(D, \mathbf{X}_m^*, \mathbf{S}), \quad (3)$$

$$C_2(\mathbf{S}) = -\frac{1}{N^-} \sum_{m:L_m=0} \Upsilon(D, \mathbf{X}_m, \mathbf{S}) \quad (4)$$

where $\Omega(\cdot)$ and $\Upsilon(\cdot)$ are defined to capture detection responses in positive and negative bags, $D(\cdot)$ is detection response of the given detector and $\mathbf{X}_m^* = \{\mathbf{x}_m^{(1)*}, \mathbf{x}_m^{(2)*}, \dots, \mathbf{x}_m^{(K)*}\}$. $\mathbf{x}_m^{(k)*}$ is the instance with maximum detection response using a target signature $\mathbf{s}^{(k)}$ in the m th positive bag such that

$$\mathbf{x}_m^{(k)*} = \arg \max_{\mathbf{x}_n \in B_m^+} D(\mathbf{x}_n, \mathbf{s}^{(k)}). \quad (5)$$

1) Detectors: In this paper, two detectors, i.e., ACE and SMF, are used to produce detection response $D(\cdot)$. Other detectors are allowed in our method. The ACE and SMF detectors using target signature \mathbf{s} for input spectral \mathbf{x}_n can be, in turn, written as [41]

$$D_{\text{ACE}}(\mathbf{x}_n, \mathbf{s}) = \hat{\mathbf{s}}^T \hat{\mathbf{x}}_n, \quad D_{\text{SMF}}(\mathbf{x}_n, \mathbf{s}) = \hat{\mathbf{s}}^T \hat{\mathbf{x}}_n \quad (6)$$

$$\hat{\mathbf{x}}_n = \frac{\hat{\mathbf{x}}_n}{\|\hat{\mathbf{x}}_n\|}, \quad \hat{\mathbf{x}}_n = \mathbf{E}^{-\frac{1}{2}} \mathbf{U}^T (\mathbf{x}_n - \boldsymbol{\mu}_b) \quad (7)$$

$$\hat{\mathbf{s}} = \frac{\hat{\mathbf{s}}}{\|\hat{\mathbf{s}}\|}, \quad \hat{\mathbf{s}} = \mathbf{E}^{-\frac{1}{2}} \mathbf{U}^T \mathbf{s} \quad (8)$$

where \mathbf{U} and \mathbf{E} are the eigenvectors and eigenvalues of the background covariance matrix \mathbf{V}_b , respectively, derived from the assumption that instances in negative bags (background) follow the normal distribution with mean vector $\boldsymbol{\mu}_b$ and covariance matrix \mathbf{V}_b .

2) Global Detection Statistics in Positive Bags: The term $\Omega(D, \mathbf{X}_m^*, \mathbf{S})$ in (3) denotes global detection statistics in the sense of using all multiple signatures in a positive bag. It should be formulated to make the learned target characterizations highlight the possible positive instances to obtain

high values for both the global and local detection statistics (in the sense of using a single concept). To obtain these abilities, $\Omega(D, \mathbf{X}_m^*, \mathbf{S})$ for ACE is defined as the average of the maximum detection responses using multiple signatures in the m th positive bag

$$\Omega_{\text{ACE}}(D, \mathbf{X}_m^*, \mathbf{S}) = \frac{1}{K} \sum_k \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{x}}_m^{(k)*}. \quad (9)$$

The monotonicity of average operator guarantees the learned signatures by (2) of the ability to maximize detection responses. It should be mentioned that the proposed method does not assume that every positive bag covers all target types. It is possible that the m th positive bag does not contain positive instances of the k th target type, and thus the maximum detection response $\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{x}}_m^{(k)*}$ could be very small. Combined with (2) and (9), this demonstrates a reasonable fact that the target types appearing in fewer positive bags should have smaller weights in objective function.

3) *Global Detection Statistics in Negative Bags*: The global detection statistics term $\Upsilon(D, \mathbf{X}_m, \mathbf{S})$ over negative instances should be small and thus obtains the ability to suppress background. It is defined for ACE as

$$\Upsilon_{\text{ACE}}(D, \mathbf{X}_m, \mathbf{S}) = \frac{1}{N_m^-} \sum_{\mathbf{x}_n \in B_m^-} \max_k \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{x}}_n. \quad (10)$$

The definition means that if maximum responses of multiple signatures \mathbf{S} over \mathbf{x}_n is minimized, all the local detection responses will have small values. Therefore, the learned concepts can effectively discriminate negative instances.

B. Diverse Multiple Characterization MIL

To explicitly encourage diversity in the learned multiple target characterizations, we augment the objective function to penalize the resulting signatures that lack diversity with respect to each other. This encourages different characterizations to detect one type of potential targets. Formally, the augmented objective function is formulated as

$$C^{+div}(\mathbf{S}) = C_1(\mathbf{S}) + C_2(\mathbf{S}) + \alpha C^{\text{div}}(\mathbf{S}) \quad (11)$$

where $C^{\text{div}}(\mathbf{S})$ is the signature diversity promoting augment, and $\alpha > 0$ is a constant that controls tradeoff between task objective and diversity.

1) *Diversity Promoting Function*: To define a feasible diversity promoting cost for the learned signatures, we formally describe diversity as how different each concept $\hat{\mathbf{s}}^{(k)}$ is from the others. There are many ways to measure the difference between concepts $\hat{\mathbf{s}}^{(k)}$ and $\hat{\mathbf{s}}^{(l)}$. This paper defines a angle-based difference measure, that is,

$$C^{\text{div}}(\mathbf{S}) = -\frac{2}{K(K-1)} \sum_{k,l,k \neq l} \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(l)}. \quad (12)$$

In (12), $\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(l)}$ is actually a cosine similarity. Thus, a larger $C^{\text{div}}(\mathbf{S})$, a larger angel between $\hat{\mathbf{s}}^{(k)}$ and $\hat{\mathbf{s}}^{(l)}$, indicates more diverse signatures. In fact, other diversity promoting functions can also be naturally used in our method only if they are differentiable from the view of easily computable [20], [43], [44].

2) *Normalization Constraint*: In addition, the constraint $\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(k)} = 1, k = 1, 2, \dots, K$ is further introduced into the diversity promoting objective function through the Lagrange method. The objective function is then written as

$$C^{+div}(\mathbf{S}) = C_1(\mathbf{S}) + C_2(\mathbf{S}) + \alpha C^{\text{div}}(\mathbf{S}) + \lambda C^{\text{con}}(\mathbf{S}) \quad (13)$$

where the constraint term $C^{\text{con}}(\mathbf{S})$ is

$$C^{\text{con}}(\mathbf{S}) = -\frac{1}{K} \sum_k |\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(k)} - 1|. \quad (14)$$

The constraint, $\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(k)} = 1, k = 1, 2, \dots, K$, is derived from the fact that $\hat{\mathbf{s}} = (\hat{\mathbf{s}}/\|\hat{\mathbf{s}}\|)$. The introduction of this constraint into the maximization framework aids in the prevention of values of $\hat{\mathbf{s}}^{(k)}$ from being arbitrarily large, which could make the first term in (13) have extreme value and erase the effects of other terms. The parameter $\lambda > 0$ is a constant that weights the constraint in objective function.

3) *Final Objective Function*: Using (3)–(14), the final objective function of the ACE specific multiple diverse hyperspectral target characterization can be finally written as

$$\begin{aligned} C_{\text{ACE}}^{+div} = & \frac{1}{N^+} \sum_{m:L_m=1} \sum_k \frac{1}{K} \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{x}}_m^{(k)*} \\ & - \frac{1}{N^-} \sum_{m:L_m=0} \frac{1}{N_m^-} \sum_{\mathbf{x}_n \in B_m^-} \max_k \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{x}}_n \\ & - \frac{2\alpha}{K(K-1)} \sum_{k,l,k \neq l} \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(l)} \\ & - \frac{\lambda}{K} \sum_k |\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(k)} - 1|. \end{aligned} \quad (15)$$

The objective for SMF can be similarly derived. As noted in (6), the difference between objectives corresponding to ACE and SMF is the use of $\hat{\mathbf{x}}_n$ and $\hat{\mathbf{x}}_m^{(k)*}$ instead of $\hat{\mathbf{x}}_n$ and $\hat{\mathbf{x}}_m^{(k)}$. In the remaining contents, the proposed multiple instance learning of multiple diverse characterizations for SMF (MILMD-SMF) and ACE detector (MILMD-ACE) are abbreviated.

IV. OPTIMIZATION

A. Gradients

The general gradient-based optimization algorithms can be used to maximize objective function (15). An important observation is that the second and fourth term in (15) is not differentiable since $\max\{\bullet\}$ and $|\bullet|$ operator is not differentiable at zero. In this paper, the noisy-or (NOR) differentiable function [36] is used to approximate $\max\{\bullet\}$ operator, and a stable subgradient method [38] is performed to compute gradient of $|\bullet|$ operator. The final gradient for MILMD-ACE can be computed as

$$\frac{\partial C_{\text{ACE}}^{+div}(\mathbf{S})}{\partial \hat{\mathbf{s}}^{(k)}} = \frac{\partial C_1}{\partial \hat{\mathbf{s}}^{(k)}} + \frac{\partial C_2}{\partial \hat{\mathbf{s}}^{(k)}} + \alpha \frac{\partial C^{\text{div}}}{\partial \hat{\mathbf{s}}^{(k)}} + \lambda \frac{\partial C^{\text{con}}}{\partial \hat{\mathbf{s}}^{(k)}}. \quad (16)$$

Then, the gradients of four terms can be computed as follows.

Computing $(\partial C_1(\mathbf{S})/\partial \hat{\mathbf{s}}^{(k)})$ and $(\partial C^{\text{div}}(\mathbf{S})/\partial \hat{\mathbf{s}}^{(k)})$: The gradients can be directly computed as

$$\frac{\partial C_1(\mathbf{S})}{\partial \hat{\mathbf{s}}^{(k)}} = \frac{1}{N^+} \sum_{m:L_m=1} \frac{1}{K} \sum_k \hat{\mathbf{x}}_m^{(k)*} \quad (17)$$

which is average of the instances with maximum detection characteristics using the learned target signatures, and

$$\frac{\partial C^{\text{div}}(\mathbf{S})}{\partial \hat{\mathbf{s}}^{(k)}} = -\frac{2}{K(K-1)} \sum_{k,l,k \neq l} \hat{\mathbf{s}}^{(l)}. \quad (18)$$

Computing $(\partial C_2(\mathbf{S})/\partial \hat{\mathbf{s}}^{(k)})$: In this paper, we use the NOR function to approximate the $\max\{\cdot\}$ operator [36]. The method approximates $\max\{\cdot\}$ over $\{z_1, z_2, \dots, z_K\}$ by a differentiable function

$$g_k(\mathbf{z}_k) = 1 - \prod_k (1 - \mathbf{z}_k) \approx \max_k \mathbf{z}_k \quad (19)$$

only if $\forall k, \mathbf{z}_k \in [0, 1]$. The corresponding derivative is

$$\frac{\partial g_k(\mathbf{z}_k)}{\partial \mathbf{z}_k} = \prod_{l \neq k} (1 - \mathbf{z}_l). \quad (20)$$

We equivalently rewrite the second term as

$$C_2(\mathbf{S}) = \frac{-1}{N^-} \sum_{m:L_m=0} \frac{1}{N_m^-} \sum_{x_n \in B_m^-} \left(2 \max_k \frac{\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{x}}_n + 1}{2} - 1 \right). \quad (21)$$

By defining $\mathbf{z}_k = ((\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{x}}_n + 1)/2) \in [0, 1]$ and using the NOR approximation of the $\max\{\bullet\}$ operator, we can obtain the approximation as

$$C_2(\mathbf{S}) = -\frac{1}{N^-} \sum_{m:L_m=0} \frac{1}{N_m^-} \sum_{x_n \in B_m^-} (2g_k(\mathbf{z}_k) - 1). \quad (22)$$

Using (19) and (20), the corresponding gradient is computed as

$$\begin{aligned} \frac{\partial C_2(\mathbf{S})}{\partial \hat{\mathbf{s}}^{(k)}} &= -2 \frac{1}{N^-} \sum_{m:L_m=0} \frac{1}{N_m^-} \sum_{x_n \in B_m^-} \frac{\partial g_k(\mathbf{z}_k)}{\partial \mathbf{z}_k} \frac{\partial \mathbf{z}_k}{\partial \hat{\mathbf{s}}^{(k)}} \\ &= -\frac{1}{N^-} \sum_{m:L_m=0} \frac{1}{N_m^-} \sum_{x_n \in B_m^-} \hat{\mathbf{x}}_n \prod_{l \neq k} (1 - \mathbf{z}_l) \\ &= -\frac{1}{N^-} \sum_{m:L_m=0} \frac{1}{N_m^-} \sum_{x_n \in B_m^-} \hat{\mathbf{x}}_n \prod_{l \neq k} \left(\frac{1 - \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{x}}_n}{2} \right). \end{aligned} \quad (23)$$

Computing $(\partial C^{\text{con}}(\mathbf{S})/\partial \hat{\mathbf{s}}^{(k)})$: The operator $|\bullet|$ in the fourth term $C^{\text{con}}(\mathbf{S})$ is not differentiable at zero. A stable subgradient approximation method is used to compute the gradient in this paper [38]. Finally, the gradient of the fourth-term $C^{\text{con}}(\mathbf{S})$ can be written as

$$\frac{\partial C^{\text{con}}}{\partial \hat{\mathbf{s}}^{(k)}} = \begin{cases} \frac{2}{K} \hat{\mathbf{s}}^{(k)}, & \text{if } \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(k)} > 1 \text{ or } (\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(k)} = 1 \\ & \text{and } \Delta T - \frac{2\lambda}{K} \hat{\mathbf{s}}^{(k)} < 0) \\ -\frac{2}{K} \hat{\mathbf{s}}^{(k)}, & \text{if } \hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(k)} < 1 \text{ or } (\hat{\mathbf{s}}^{(k)T} \hat{\mathbf{s}}^{(k)} = 1 \\ & \text{and } \Delta T - \frac{2\lambda}{K} \hat{\mathbf{s}}^{(k)} > 0) \\ 0, & \text{else} \end{cases} \quad (24)$$

where ΔT is defined as

$$\Delta T = \frac{\partial C_1(\mathbf{S})}{\partial \hat{\mathbf{s}}^{(k)}} + \frac{\partial C_2(\mathbf{S})}{\partial \hat{\mathbf{s}}^{(k)}} + \frac{\partial C^{\text{div}}(\mathbf{S})}{\partial \hat{\mathbf{s}}^{(k)}} \quad (25)$$

Algorithm 1 Learn Joint Diverse Target Characterizations

Input: $B = \{B_1, B_2, \dots, B_M\}$, $L = \{L_1, L_2, \dots, L_M\}$, K, α, λ
Output: $\mathbf{S} = \langle \mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(K)} \rangle$

- 1: Compute μ_b and \mathbf{V}_b using instances in negative bags
- 2: Compute all $\hat{\mathbf{x}}_n$ using (7)
- 3: If the ACE specific method, compute $\hat{\mathbf{x}}_n$ using (7)
- 4: Initialize $\hat{\mathbf{S}}^{(0)}$ as instance subset in positive bags resulting in largest objective function value
- 5: **repeat**
- 6: **for** $k = 1, 2, \dots, K$ **do**
- 7: Update $\hat{\mathbf{x}}_m^{(k)*}$ for each positive bag using (5)
- 8: Update $\hat{\mathbf{s}}^{(k)}$ using gradients computed as (16) for ACE and (16) with $\hat{\mathbf{x}}_n$ and $\hat{\mathbf{x}}_m^{(k)*}$ instead of $\hat{\mathbf{x}}_n$ and $\hat{\mathbf{x}}_m^{(k)*}$ for SMF
- 9: **end for**
- 10: **until** Stopping Criterion Reached
- 11: **return** $\mathbf{S} = \langle \mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(K)} \rangle$, $\mathbf{s}^{(k)} = \mathbf{U} \mathbf{E}^{\frac{1}{2}} \hat{\mathbf{s}}^{(k)}$, $k = 1, 2, \dots, K$

Using (17), (18), (23), and (24), the final gradient can be easily computed by (16).

B. Algorithm

Under the gradient-based optimization framework, the diverse multiple target characterizations can be jointly learned. The proposed algorithm is summarized in Algorithm 1, where the details about initialization and stopping criterion are given as follows.

1) Initialization: In theory, the positive instances corresponding to different target types can be used as the initializations of target signatures. However, it is difficult to precisely identify the positive instances given only bag-level labels. The k-mean and random sample mean methods [22], [26] can also be used to extract spectral curves as initializations of multiple signatures. In this paper, $\hat{\mathbf{S}}^{(0)}$ is obtained as the instance subset (with K elements) in positive bags resulting in largest objective function value. The number of instance subset candidates is $\binom{N}{K}$, where N and K are the number of positive instances and target types, respectively. When N and K are relatively large, the computations of $\binom{N}{K}$ objective function values are timely prohibitive. To make the initialization computable, we first use the k-mean to cluster positive instances into C groups. The instance closest to the clustering center is selected as the representative of the cluster. The initialization is performed over the C representative instances, and thus only $\binom{C}{K}$ objective function values are needed to be computed. The initialization is thus significantly accelerated since C is usually much smaller than N .

For the diversity-promoting term incorporated into the objective function, the instances selected for initialization are different in the sense of the given distance metric and thus could cover different target types. Through the subsequent diversified joint learning method, the initializing instances

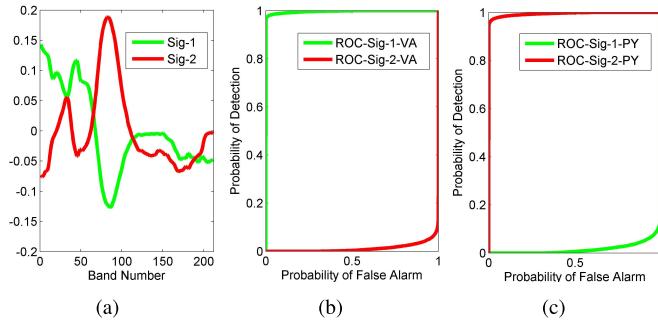


Fig. 3. Results from MILDM-ACE over the simulated hyperspectral data set. (a) Learned target signatures. (b) and (c) ROCs of ACE using learned target signatures to detect VA and PY, respectively.

possibly converge to the discriminative multiple diverse target characterizations.

2) *Stopping Criterion*: The stopping criterion of the learning method could be the maximum iteration steps or the difference between the obtained target signatures from two successive iterations. In this paper, we use both the criteria, and the iteration will be stopped when one of the criteria is reached.

V. EXPERIMENTS

In this section, the proposed multiple diverse target characterization method is evaluated over both the simulated and real-world hyperspectral data sets through using the learned target signatures for target detection. Corresponding to the interactive target detection application, the Oracle evaluation is used to evaluate the proposed method. Let testing samples and corresponding ground truth (for a given target type) be $X = \{\mathbf{x}_n, n = 1, 2, \dots, N\}$ and $Y = \{y_n, n = 1, 2, \dots, N\}$, respectively. $\hat{Y}^{(k)} = \{\hat{y}_n^{(k)}, n = 1, 2, \dots, N\}$ denotes target detection result of the learned target signature $s^{(k)}$. The Oracle evaluation is

$$\text{OE} = \max_k R(Y, \hat{Y}^{(k)}) \quad (26)$$

where $R(Y, \hat{Y}^{(k)})$ is an usual evaluation for the single output, and the area under the curve (AUC) is used in this paper. Therefore, the Oracle metric defined in (26) is actually the highest score over all outputs. In fact, the Oracle evaluation has been extensively used in many prior works to evaluate multiple-prediction systems [8], [27], [30]. More applications and explanations of the Oracle evaluation can be found in [30].

A. Results Over Simulated Data Set

1) *Experimental Data Set*: The method used to simulate hyperspectral data is similar to that in [41]. The main difference is that the positive bags are composed of multiple types of targets. A linear mixing model and several spectra selected from the ASTER spectral library [3] are used to simulate the hyperspectral data. In this paper, we use Balsalt, Red Slate, and Phyllite spectra as background endmembers and Pyroxenite (PY) and Verde Antique (VA) as target endmembers. The simulated hyperspectral data have 211 bands, and the testing

and training data were generated separately. The generated testing data set is composed of 25 000 true negative and 25 000 true positive instances with an average target proportion value of 0.15. The training data set contains 3 positive bags and 47 negative bags, with each bag having 20 instances. A positive bag has two positive instances for each target type.

The reason that positive bags contain only a small number of positive instances is to simulate the usual situations, where objects appear as small or subpixel targets in hyperspectral imagery. Moreover, the fewer positive instances make the signature learning more challenging and thus can comprehensively evaluate the proposed method. In addition, zero-mean Gaussian noise was added to the simulated training and testing data such that SNR was 20 dB.

In fact, the parameters in the experimental setting on the number of true target signatures in the positive bag, the proportion/mixing level of these target signatures in the positive bags and the SNR of the target signature within noise virtually affect the performance of the proposed method. They were experimentally analyzed in work [41]. Since our method is basically derived from work [41], it should behave similarly with the method in [41] when these parameters change. For page limitation, we do not present the similar results, which can be found in [41].

2) *General Performance*: At first, we present a brief overview of the proposed method's merits. In this set of experiments, the weight λ of the normalization term is fixed as constant 1, and the step η is fixed as 0.01. Other two parameters (K and α) were set as $K = 2$ and $\alpha = 1$, respectively. The effects of these parameters on the method's performance will be evaluated subsequently. We run the proposed method over the machine with a 3.40-GHz Intel (R) Core (IM) i7 and 32 GB memory. The proposed method implemented by our unoptimized MATLAB code took about 15 s.

Fig. 3 shows the learned target signatures and the receiver operating characteristic (ROC) curves of the learned signatures for the given target types. The learned signatures accurately represent the targets [see Fig. 3(a)] and can effectively discriminate the object from the background [see Fig. 3(b) and (c)], where one of the AUCs computed from the ROCs is approximately 1]. Moreover, the learned signatures can effectively discriminate each other and focus on only one type of target [see Fig. 3(b) and (c), where the AUC of one detection curve tends to be 1 and the other is close to 0]. The discrimination between the learned signatures is mainly derived from the introduction of signature diversity into the learning process. More results from ten runs of the experiments show similar behaviors. Finally, the proposed MILMD-SMF obtained the average Oracle of ten runs as 0.9894 and 0.9808 when detect PY and VA, respectively, while the MILMD-ACE obtained similar results as 0.9865 and 0.9764.

3) *Comparisons With Other Methods*: The results in the proposed MILMD-SMF and MILMD-ACE were compared with the recent characterization methods: MI-HE [22], [26], eFUMI [24], MIL-SMF [41], MIL-ACE [41], and EM-DD [42]. eFUMI was initialized as outlined in [41]. EM-DD estimates both the point and scale value and has the same parameter setting as that in [41]. Table I lists the

TABLE I
RESULTS OF DIFFERENT METHODS FOR TARGET PY AND VA. THREE
BEST RESULTS ARE IN BOLD

Methods	PY	VA
MILMD-SMF	0.9894 ± 0.0086	0.9808 ± 0.0055
MILMD-ACE	0.9865 ± 0.0105	0.9764 ± 0.0064
MI-HE	0.9475 ± 0.0862	0.9357 ± 0.0836
MI-HE-Oracle	0.9895 ± 0.0085	0.9812 ± 0.0053
eFUMI	0.8107 ± 0.0959	0.6492 ± 0.1112
eFUMI-Oracle	0.8903 ± 0.1947	0.8825 ± 0.1992
MIL-SMF	0.8839 ± 0.2763	0.5936 ± 0.2191
MIL-ACE	0.8660 ± 0.2933	0.5375 ± 0.2464
EM-DD	0.4685 ± 0.0325	0.4595 ± 0.0955

experimental results of the average values of ten runs of experiments for each method. The proposed methods obtained the best results when detecting PY and VA, where EM-DD obtained the worst results. MIL-SMF and MIL-ACE produced one target signature that combines the useful information from different target types together, which cannot yield a good performance for even one type of target. eFUMI can simultaneously learn a target signature and multiple background signatures. The learned target signatures yielded values of 0.8107 and 0.6492 for the PY and VA, respectively. The newly proposed MI-HE worked very well on learning multiple target signatures simultaneously. MI-HE can learn a set of target signatures and also background signatures. Using the learned target signatures directly to detect PY and VA obtained 0.9475 and 0.9357 detection AUCs.

Further analysis of the learned signatures by eFUMI and MI-HE find that eFUMI and MI-HE could treat the targets as background endmembers, i.e., the accurate target signatures could lie in the learned background signature set. Therefore, to comprehensively evaluate the performance of eFUMI and MI-HE, we further evaluate their learned background signatures and compute the Oracle evaluations, where all background signatures and target signatures are used to detect the given object type and the best results are deemed as the method's performance. The corresponding results are labeled as eFUMI-Oracle and MI-HE-Oracle. Under this evaluation framework, the eFUMI-Oracle values for PY and VA are 0.8903 and 0.8825, and MI-HE-Oracle values are 0.9895 and 0.9812. Our proposed method is still better than eFUMI and performed similarly with MI-HE. To comprehensively evaluate the proposed method, we further analyze the effects of important parameters on the performance.

4) *Effects of K*: Target signature number K determines the possibility of learning representations of all actual target types presented in training bags. In theory, all the target signatures could be learned when K is no less than the number of actual object types. Fig. 4 shows the detection results using target signatures learned with different K values. The detection performance increases with a larger K and is stable when K is larger than 2, which is the number of actual target types in training bags.

5) *Effects of α* : Another important parameter in the proposed method is α , which controls the diversity of the learned target signatures. Examples of the learned target signatures

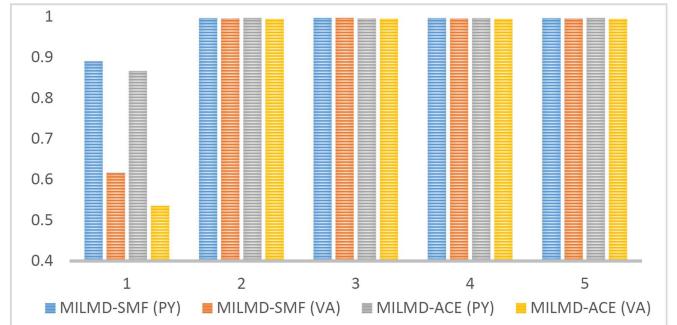


Fig. 4. Effects of the concept number parameter K on detection results. Horizontal axis: K . Vertical axis: detection AUCs.

with different levels of α are shown in Fig. 5. The proposed method learned two highly similar signatures with a very small α ($\alpha = 0.001$). With the increase in α , the learned signatures become notably different. When α is larger than 10, the two learned representations tend to be inverse signatures, meaning that they are most different. The diversity promoting term [(12)] in objective function without weight α is computed to give a quantitative evaluation of the signature diversity. Fig. 6(a) shows the average diversity of ten runs of the experiments for each different α . It can be noted that the increases in parameter α lead to increased diversity. Fig. 6(b) further illustrates the detection performance using the signatures learned with different levels of α . The Oracle evaluation increased with increases in α initially, and then the performance decreased when the value of α is larger than 1. The reason can be observed in Fig. 5. The learned signatures tend to guarantee fitting actual target signatures through the first two terms in the learning objective [(13)]. However, with increases in α , the proposed method will emphasize signatures' diversity, making the learned signatures far from actual signatures and thus resulting in decreased detection performance.

6) *Effects of Target Signature Number*: At the end of this section, the simulated data set with multiple target types are used to demonstrate the performance as the number of target signatures that are needed to be estimated is increased. To simulate this situation, more target endmembers are extracted from ASTER spectral library, including Porphyritic Biotite Granite, Dolomite Marble (DM), Limestone Conglomerate-Breccia, and Calcareous Shale. Each set of experiments add one more target endmember to produce positive instances in the positive bags. Other experimental settings to produce the simulated data sets are the same as the previous experiments. Through this experimental setting, we will obtain the data sets with 2–6 target signatures needed to be estimated. We performed ten runs of data simulation and learning methods and computed the average performance for each target signature number. Fig. 7 shows results with the increase in target signature number. The detection performance decreased for all target types when new target types were added into the training samples. It follows our expectation since the newly added target types make the data and also the discrimination of

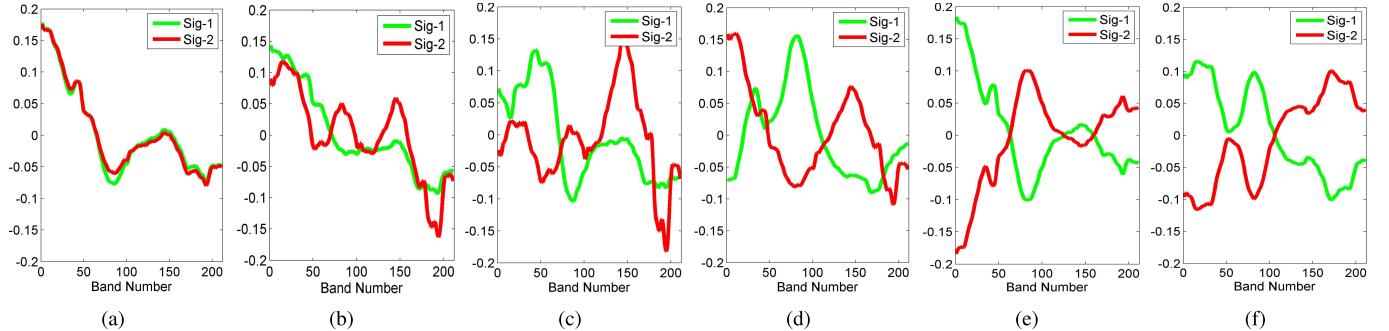


Fig. 5. Learned target signatures using different values of α . (a)–(f) Values are 0.001, 0.01, 0.1, 1, 10, and 100, respectively.

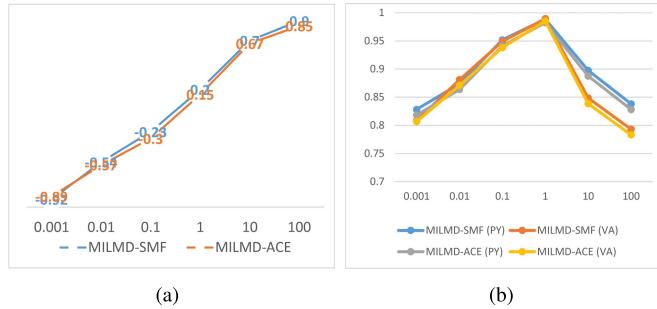


Fig. 6. Effects of parameter α on (a)signature diversity and (b) detection performance, where horizontal axes show the values of parameter α and vertical axis of (a) shows the diversity evaluations and (b) the values of detection AUC.

multiple target signatures more complex, especially when the target types are similar.

B. Results Over Real-World Data Set

1) *Experimental Data Sets*: Two real-world data sets, i.e., MUUFL Gulfport [15] and SHARE 2012 AVON [16], were used to test performances of the proposed methods. The MUUFL Gulfport hyperspectral image has 325×337 pixels with 72 bands and 1-m spatial resolution. The data set contains two sets of data (named Flight 1 and Flight 3) from two flights over same scene at the same altitude and with the same spatial resolution, but at different times. The imaging scene is shown in Fig. 8(a). More details on the data set can be found in [41]. The SHARE AVON hyperspectral data set collected near Rochester, NY, USA, has 388×556 pixels with 360 bands and 1-m spatial resolution. The data set contains two sets of data: 1) AVON AM collected in the morning of September 20, 2012 and 2) AVON PM collected in the afternoon of the same day. The imaging scene is shown in Fig. 9(a). More details on the data set can be found in [16].

2) *Real-World Application Tasks*: In the MUUFL Gulfport data set, four types of man-made targets are scattered in the imaged scene. They are cloth panels of four different colors: brown (15 examples), dark green (15 examples), faux vineyard green (FVG) (12 examples), and pea green (15 examples). Each type of target varies in size among 0.5, 1 and 3 m. With the 1-m spatial resolution, a target covers only several pixels or a subpixel. Fig. 8(b) shows their spatial locations

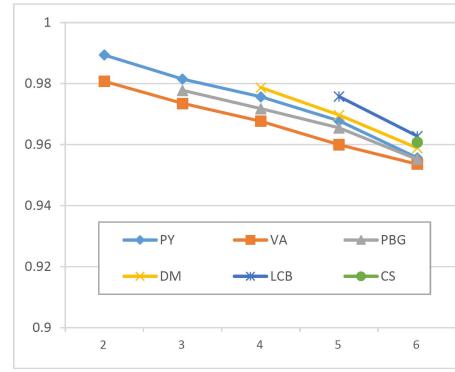


Fig. 7. Effects of number of target signatures needed to be learned on the detection performance. The horizontal axis shows the number of target signatures and vertical axis shows the detection AUCs.

using different colors for different target types. Many of the targets are partially or fully occluded by live oak trees. Therefore, it is difficult to identify all the targets to construct ideal positive bags such as that used in [41] for MIL target characterization. In this paper, we consider feasible real-world applications, which follow the flowchart shown in Fig. 1. We first select the suspicious regions as positive bags, and then randomly select ten regions with random sites and sizes as negative bags. Fig. 8(c) and (d) shows two individual tasks with the selected training bags. Just as shown in the figures, it is very possible that a selected positive bag contains multiple target types, such as the dark green and FVG targets in Fig. 8(c).

In the SHARE AVON data set, two types of man-made targets are scattered in the imaged scene. They are tarps of two different colors: brown (12 examples) and blue (12 examples). Fig. 9(b) shows their spatial locations using different colors for different target types. The selected positive bag and negative bag following the real-world application flowchart are shown in Fig. 9(c). Similar with situation in the MUUFL Gulfport data set, the selected positive bag could contain multiple target types.

3) *Learned Target Signatures*: The proposed MILMD-SMF and MILMD-ACE were run with $\alpha = 1$ over two real-world data sets. Because the selected positive bags mainly contain two types of targets, we set parameter K as 3 to guarantee effectiveness and efficiency simultaneously. Fig. 10(a) and (b)

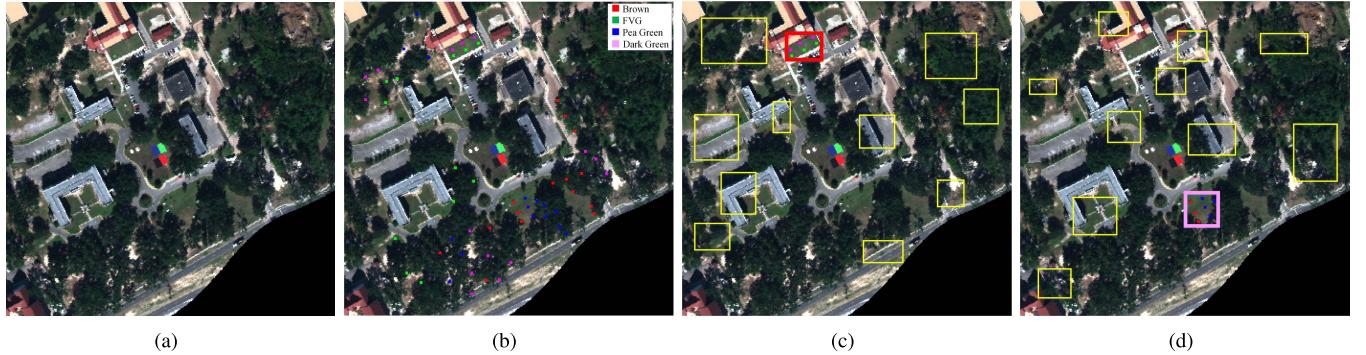


Fig. 8. MUUFL Gulfport hyperspectral data set and two tasks using it. (a) RGB image of imaging scene. (b) True sites of four types of targets (labeled with different colors). (c) Selected positive bag (red box) and negative bags (yellow boxes) in task 1. (d) Selected positive bag (pink box) and negative bags (yellow boxes) in task 2.

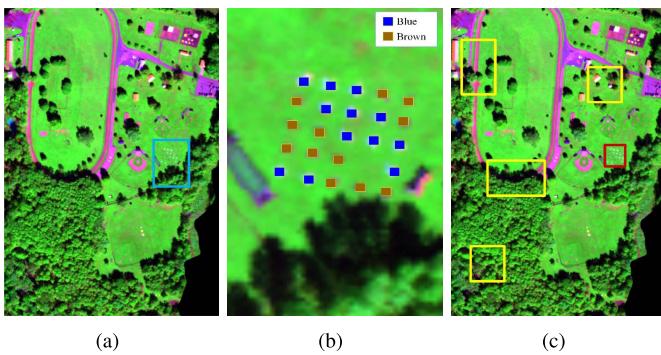


Fig. 9. SHARE AVON hyperspectral data set and the task using it. (a) RGB image of imaging scene. (b) True sites of two types of targets (labeled with different colors). (c) Selected positive bag (red box) and negative bags (yellow boxes).

shows the learned multiple target signatures for tasks 1 and 2, respectively, over the MUUFL Gulfport data set. The results in the proposed MILMD-SMF and MILMD-ACE over the SHARE AVON data set are presented in Fig. 11(a) and (b), respectively. It can be noted that the proposed methods can effectively learn representations of the considered targets. Moreover, the learned diverse signatures can discriminate a target type from both the background endmembers and other target types. In addition, as auxiliary products, some learned signatures are highly different from target signatures (such as the first signature from MILMD-SMF and the third signature from MILMD-ACE in task 1 over the MUUFL Gulfport data set), but still may denote some type of potential targets, which provides us with the opportunity to find potentially new target types.

4) Target Detection Results: For the MUUFL Gulfport data set, the detection performance was evaluated by Normalized AUC (NAUC), where the area was normalized out to a false alarm rate (FAR) of 1×10^{-3} falsealarms/m² [17]. Fig. 12 shows the detection responses of tasks 1 and 2, where the proposed methods were trained on Flight 3 and tested on Flight 1. Because MILMD-SMF and MILMD-ACE obtained similar results, only the results in MILMD-ACE are presented for conciseness. To clearly show detection performance, the true spatial position of targets and NAUCs were overlaid

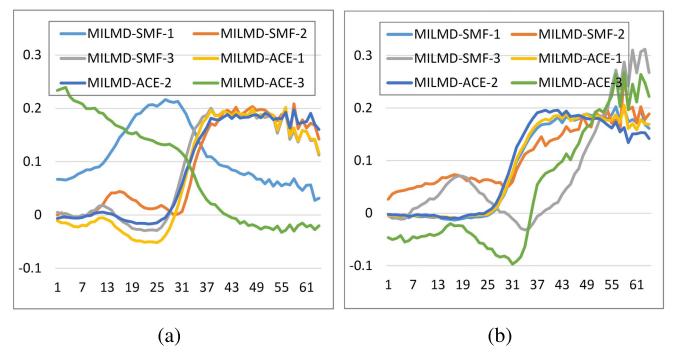


Fig. 10. Learned multiple diverse target signatures by the proposed methods in (a) task 1 and (b) task 2 over the MUUFL Gulfport data set. MILMD-SMF-*k* and MILMD-ACE-*k* mean the *k*th signature learned by MILMD-SMF and MILMD-ACE, respectively.

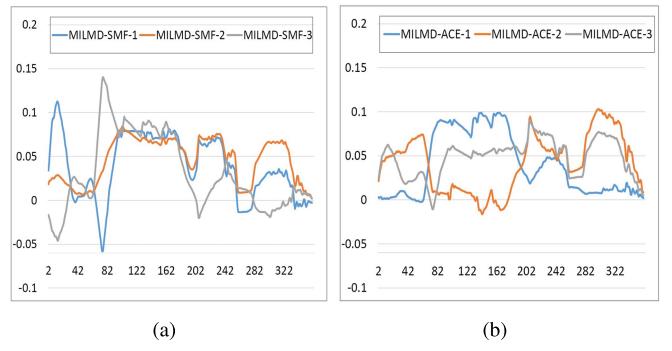


Fig. 11. Learned multiple diverse target signatures over the SHARE AVON data set by the proposed (a) MILMD-SMF and (b) MILMD-ACE, respectively. MILMD-SMF-*k* and MILMD-ACE-*k* mean the *k*th signature learned by MILMD-SMF and MILMD-ACE, respectively.

on the detection results as colored boxes. The colors' semantic meanings are same as that in Fig. 8(b).

Fig. 12 demonstrates that the learned signatures produced diverse outputs and each signature focused on detecting specific individual target type. As an example, we used the learning results of task 1 to detect FVG and dark green. The first learned signature effectively detected FVG but failed to find dark green [see Fig. 12(a), where the NAUC is 0.5963 for FVG versus 0 for dark green], while the second learned

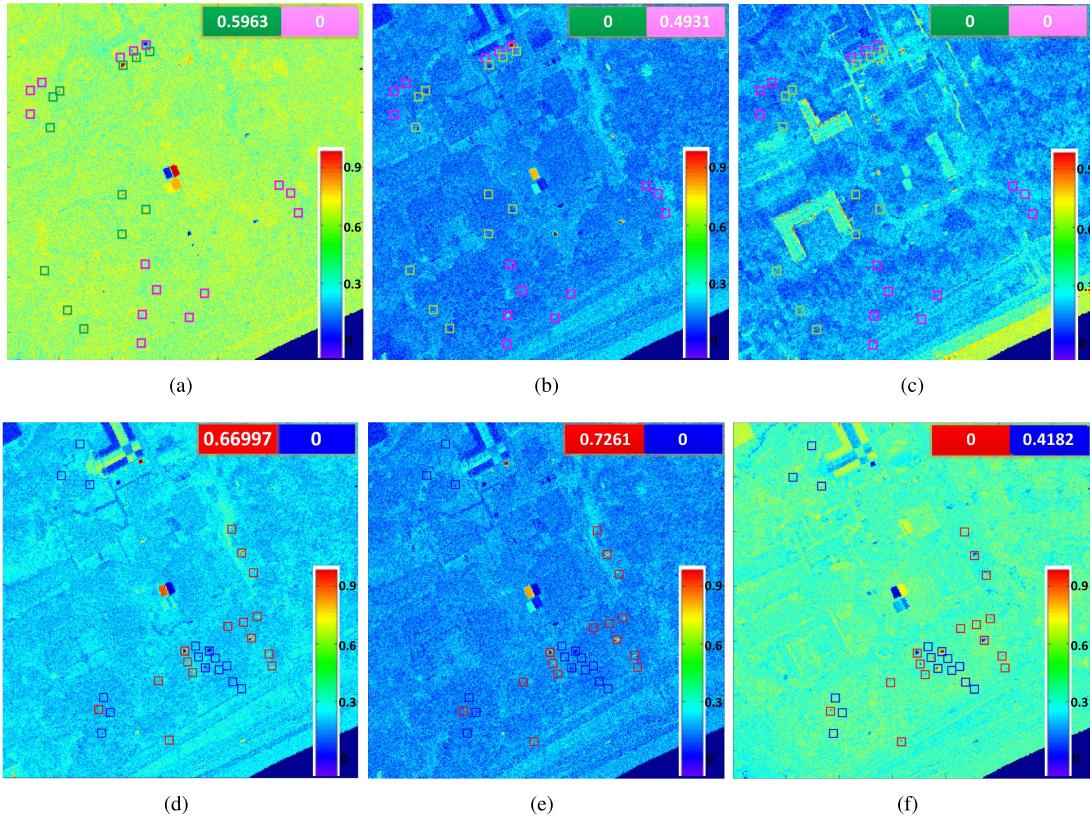


Fig. 12. Detection confidences and AUCs using the learned multiple target signatures by MILMD-ACE in (a)–(c) task 1 and (d)–(f) task 2 (d–f) over the MUUFL Gulfport data set. The colormap is overlaid in each result. It is better to zoomed-in view to see the details.

signature effectively detected the majority of the dark green targets but lost all FVG [see Fig. 12(b), where the NAUC is 0 for FVG versus 0.4931 for dark green]. Fig. 12(c) further demonstrated that although the third learned signature did not find the considered targets (with an NAUC 0 for both target types), it still effectively suppressed the background and found new meaningful materials. Here, the proposed method learned a representation for cement-like materials.

Fig. 13 shows the detection responses over the SHARE AVON data set. The proposed methods were trained on AVON AM and tested on AVON PM. As can be seen, the detection results match the RGB image displaying the locations of the targets very well. Moreover, Fig. 13 demonstrates that the learned signatures produced diverse outputs and focused individually on detecting specific target type.

5) Comparisons With Other Methods: The recent characterization methods including MI-HE, MI-SMF, MI-ACE, eFUMI, and EM-DD are used for the performance comparisons. The results over the MUUFL Gulfport data set are presented in Tables II and III. In both tasks, MI-SMF and MI-ACE obtained effective target characterizations for only one kind of target type. EM-DD cannot fully extract effective representations for the considered target types. Because the initialization of eFUMI will affect results, the average NAUC of ten runs of experiments is used to evaluate performance. Tables II and III demonstrate that the target signatures obtained by eFUMI yielded effective characterizations for only the brown target (with NAUCs as 0.479 and 0.729 in two dif-

TABLE II
RESULTS OF DIFFERENT METHODS OVER THE MUUFL GULFPORT DATA
SET: TRAIN ON FLIGHT 1, TEST ON FLIGHT 3. THREE
BEST RESULTS ARE IN **BOLD**

Methods	Task 1		Task 2	
	FVG	DG	PG	BR
MILMD-SMF	0.578	0.391	0.330	0.451
MILMD-ACE	0.486	0.395	0.331	0.482
MI-HE	0.579	0.092	0.335	0.086
MI-HE-Oracle	0.579	0.390	0.335	0.450
MI-SMF	0.0	0.388	0.0	0.446
MI-ACE	0.007	0.097	0.0	0.035
eFUMI	0.057	0.036	0.041	0.479
eFUMI-Oracle	0.357	0.329	0.279	0.479
EM-DD	0.0	0.0	0.0	0.0

ferent experimental settings, respectively). Because eFUMI also produces multiple signatures for background endmembers, useful information could be captured in the background signatures. Thus, we further treat background signatures as the possible target signatures and computed the Oracle for eFUMI (eFUMI-Oracle). The eFUMI-Oracle yielded better results than eFUMI, MIL-SMF, and MIL-ACE overall. However, our proposed methods still obtained the best results for the majority of target types. The recently proposed MI-HE also focuses on learning multiple target signatures simultaneously. The results show that MI-HE can learn accurate signatures for targets FVG and PG. Similar with the eFUMI-Oracle, the Oracle for MI-HE (MI-HE-Oracle) is further computed

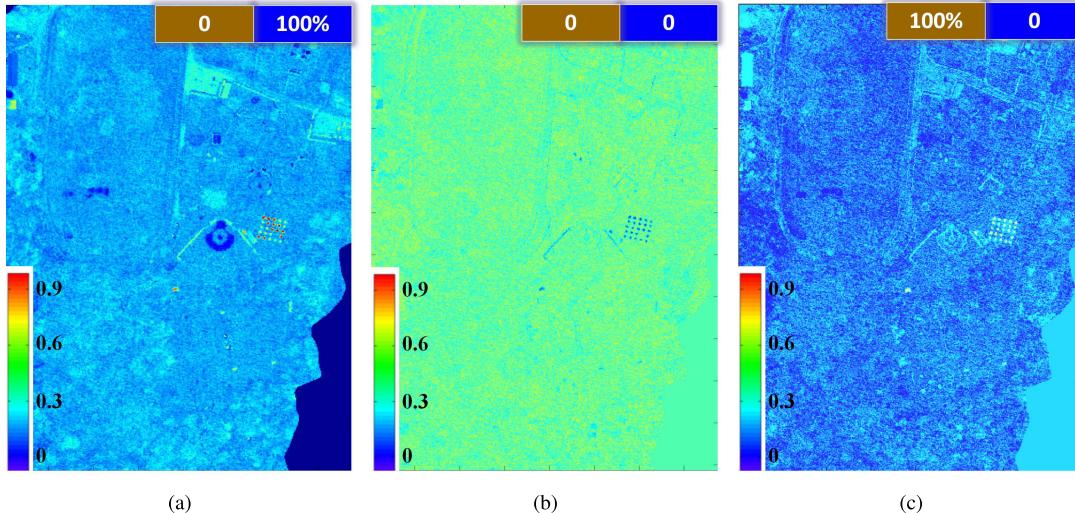


Fig. 13. Detection confidences using the learned three multiple target signatures by MILMD-ACE over the SHARE AVON data set. The colormap is overlaid in each result. It is better to zoomed-in view to see the details.

TABLE III

RESULTS OF DIFFERENT METHODS OVER THE MUUFL GULFPORT DATA SET: TRAIN ON FLIGHT 3, TEST ON FLIGHT 1. THREE BEST RESULTS ARE IN BOLD

Methods	Task 1		Task 2	
	FVG	DG	PG	BR
MILMD-SMF	0.650	0.517	0.437	0.735
MILMD-ACE	0.596	0.493	0.418	0.726
MI-HE	0.651	0.125	0.086	0.736
MI-HE-Oracle	0.651	0.510	0.351	0.736
MI-SMF	0.0	0.4315	0.0	0.719
MI-ACE	0.015	0.195	0.0	0.103
eFUMI	0.058	0.073	0.013	0.729
eFUMI-Oracle	0.487	0.412	0.297	0.729
EM-DD	0.0	0.0	0.0	0.012

TABLE V

RESULTS OF DIFFERENT METHODS OVER THE MUUFL GULFPORT DATA SET FOR THE MERGED TASK: TRAIN ON FLIGHT 3, TEST ON FLIGHT 1. THREE BEST RESULTS ARE IN BOLD

Methods	FVG	DG	PG	BR
MILMD-SMF	0.611	0.486	0.401	0.683
MILMD-ACE	0.556	0.461	0.392	0.675
MI-HE	0.620	0.119	0.090	0.684
MI-HE-Oracle	0.620	0.475	0.399	0.684
MI-SMF	0.0	0.4315	0.0	0.676
MI-ACE	0.043	0.179	0.056	0.115
eFUMI	0.089	0.053	0.027	0.717
eFUMI-Oracle	0.478	0.409	0.268	0.717
EM-DD	0.0	0.0	0.023	0.017

TABLE IV

RESULTS OF DIFFERENT METHODS OVER MUUFL GULFPORT DATA SET FOR THE MERGED TASK: TRAIN ON FLIGHT 1, TEST ON FLIGHT 3. THREE BEST RESULTS ARE IN BOLD

Methods	FVG	DG	PG	BR
MILMD-SMF	0.510	0.361	0.301	0.410
MILMD-ACE	0.429	0.364	0.300	0.438
MI-HE	0.505	0.082	0.302	0.091
MI-HE-Oracle	0.505	0.362	0.302	0.411
MI-SMF	0.479	0.102	0.0	0.101
MI-ACE	0.092	0.071	0.0	0.010
eFUMI	0.068	0.086	0.082	0.475
eFUMI-Oracle	0.348	0.327	0.287	0.475
EM-DD	0.0	0.0	0.0	0.0

to comprehensively evaluate the performance of MI-HE. The better MI-HE-Oracle evaluations compared with MI-HE evaluations demonstrate that the learned background signatures from MI-HE could be the accurate target characterizations. In addition, even compared with MI-HE-Oracle evaluations, similar results can be observed in Tables II and III from our proposed method.

Although tasks 1 and 2 demonstrate the real-world applications where the users usually focus on only one target type

(but with other target types inevitably mixed in the selected training samples), more experiments merging two positive bags (selected from both tasks 1 and 2) to learn all four signatures simultaneously were performed to give comprehensive evaluation of the proposed method. Tables IV and V show that with increase in the number of target types needed to be learned, the proposed method can still obtain satisfactory results although there is a little NAUC decrease for each target type. Similar with the previous results, MI-HE can also produce accurate characterizations using the learned signatures for targets and background (see the NAUCs of MI-HE-Oracle). Moreover, our proposed method and MI-HE performed significantly better than other methods.

Over the SHARE AVON data set, the ROC curve analysis was conducted using cross-validation between AVON AM and AVON PM. The ROC curves were generated using the Bullwinkle Scoring algorithm [17]. Since SHARE AVON data set did not contain exact pixel-level labels, it is difficult to compute the NAUC to evaluate the methods. Therefore, the probability of detection (PD) and FAR are used to evaluate the methods over the SHARE AVON data set. The detailed values of PD versus FAR is shown in Tables VI and VII, where the FAR is in the unit of false alarms /m². Similar with

TABLE VI
RESULTS OF DIFFERENT METHODS OVER THE SHARE AVON DATA SET: TRAIN ON AVON AM, TEST ON AVON PM

Methods	FAR: 1×10^{-3}		FAR: 1×10^{-4}		FAR: 1×10^{-5}	
	BLUE	BROWN	BLUE	BROWN	BLUE	BROWN
DMMI-SMF	100%	100%	100%	100%	100%	100%
DMMI-ACE	100%	100%	100%	100%	100%	100%
MI-HE	100%	16.67%	100%	16.67%	100%	8.33%
MI-HE-Oracle	100%	100%	100%	100.0%	100%	100.0%
MI-SMF	100%	16.67%	100%	0.0%	100%	0.0%
MI-ACE	100%	16.67%	100%	0.0%	100%	0.0%
eFUMI	100%	8.33%	100%	0.0%	100%	0.0%
eFUMI-Oracle	91.67%	100%	75%	100%	100%	41.67%
EM-DD	25.00%	0.0%	0.0%	0.0%	0.0%	0.0%

TABLE VII
RESULTS OF DIFFERENT METHODS OVER THE SHARE AVON DATA SET: TRAIN ON AVON PM, TEST ON AVON AM

Methods	FAR: 1×10^{-3}		FAR: 1×10^{-4}		FAR: 1×10^{-5}	
	BLUE	BROWN	BLUE	BROWN	BLUE	BROWN
DMMI-SMF	100%	100%	100%	100%	100%	100%
DMMI-ACE	100%	100%	100%	100%	100%	100%
MI-HE	100%	16.67%	100%	16.67%	100%	16.67%
MI-HE-Oracle	100%	100%	100%	100%	100%	100%
MI-SMF	100%	16.67%	100%	0.0%	100%	0.0%
MI-ACE	100%	16.67%	100%	0.0%	100%	0.0%
eFUMI	100%	8.33%	100%	0.0%	100%	0.0%
eFUMI-Oracle	83.33%	100%	58.33%	100%	100%	25.00%
EM-DD	25.00%	0.0%	0.0%	0.0%	0.0%	0.0%

the results over the MUUFL Gulfport data set, our proposed methods and the recently proposed MI-HE generally yielded better results for the given target types, especially when the FAR is low (1×10^{-5}).

VI. CONCLUSION AND DISCUSSION

This paper proposed a new MIL method to produce multiple diverse hyperspectral target signatures for the feasible real-world hyperspectral target characterization and detection tasks. The proposed method can address the problems arising from the mixed data, lack of pixel-level labels, and multiple target types in training samples.

The proposed methods were evaluated in terms of Oracle accuracy and presented multiple detection results for human interactions. It would be interesting to investigate the impact of the proposed methods on human experience. In addition, further consideration should be given to fusion methods to combine multiple detection results. Finally, introducing other diversity promoting terms into the objective and analyzing their behaviors on the results is an important future topic.

REFERENCES

- [1] N. Akhtar and A. Mian, "Nonparametric coupled Bayesian dictionary and classifier learning for hyperspectral classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4038–4050, Sep. 2018.
- [2] M. Bader-EI-Den, E. Teitei, and T. Perry, "Biased random forest for dealing with the class imbalance problem," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. doi: [10.1109/TNNLS.2018.2878400](https://doi.org/10.1109/TNNLS.2018.2878400).
- [3] A. Baldridge, S. J. Hook, C. L. Grove, and G. Rivera, "The aster spectral library version 2.0," *Remote Sens. Environ.*, vol. 113, no. 4, pp. 711–715, 2009.
- [4] M.-A. Carboneau, E. Granger, and G. Gagnon, "Bag-level aggregation for multiple-instance active learning in instance classification problems," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. doi: [10.1109/TNNLS.2018.2869164](https://doi.org/10.1109/TNNLS.2018.2869164).
- [5] Y. Chen, J. Bi, and J. Z. Wang, "MILES: Multiple-instance learning via embedded instance selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1931–1947, Dec. 2006.
- [6] V. Cheplygina, D. M. J. Tax, and M. Loog, "Dissimilarity-based ensembles for multiple instance learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1379–1391, Jun. 2016.
- [7] R. G. Cinbis, J. Verbeek, and C. Schmid, "Weakly supervised object localization with multi-fold multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 189–203, Jan. 2017.
- [8] D. Dey, V. Ramakrishna, M. Hebert, and J. A. Bagnell, "Predicting multiple structured visual interpretations," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 2947–2955.
- [9] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3862–3871.
- [10] B. Du, T. Xinyao, Z. Wang, L. Zhang, and D. Tao, "Robust graph-based semisupervised learning for noisy labeled data via maximum correntropy criterion," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1440–1453, Apr. 2019.
- [11] B. Du, S. Wang, C. Xu, N. Wang, L. Zhang, and D. Tao, "Multi-task learning for blind source separation," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4219–4231, Sep. 2018.
- [12] B. Du, M. Zhang, L. Zhang, R. Hu, and D. Tao, "PLTD: Patch-based low-rank tensor decomposition for hyperspectral images," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 67–79, Jan. 2017.
- [13] M. Fang, T. Zhou, J. Yin, Y. Wang, and D. Tao, "Data subset selection with imperfect multiple labels," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. doi: [10.1109/TNNLS.2018.2875470](https://doi.org/10.1109/TNNLS.2018.2875470).
- [14] Z. Fu, A. Robles-Kelly, and J. Zhou, "MILIS: Multiple instance learning with instance selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 958–977, May 2011.
- [15] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "Muufl gulfport hyperspectral and lidar airborne data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. 2013-570, 2013.
- [16] A. M. Giannandrea *et al.*, "The SHARE 2012 data campaign," *Proc. SPIE*, vol. 8743, p. 87430F, May 2013.
- [17] T. Glenn, A. Zare, P. Gader, and D. Dranishnikov. (2013). *Bullwinkle: Scoring Code for Sub-Pixel Targets (Version 1.0) [Software]*. [Online]. Available: <http://engineers.missouri.edu/zarea/code/>
- [18] C. Gong, D. Tao, S. J. Maybank, W. Liu, G. Kang, and J. Yang, "Multi-modal curriculum learning for semi-supervised image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3249–3260, Jul. 2016.

- [19] C. Gong, H. Zhang, J. Yang, and D. Tao, "Learning with inadequate and incorrect supervision," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2017, pp. 889–894.
- [20] Z. Gong, P. Zhong, Y. Yu, and W. Hu, "Diversity-promoting deep structural metric learning for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 371–390, Jan. 2018.
- [21] K.-j. Hsu, Y.-Y. Lin, and Y.-Y. Chuang, "Augmented multiple instance regression for inferring object contours in bounding boxes," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1722–1736, Apr. 2014.
- [22] C. Jiao, "Target concept learning from ambiguously labeled data," Ph.D. dissertation, Univ. Missouri, Columbia, MO, USA, 2017.
- [23] C. Jiao, B.-Y. Su, P. Lyons, A. Zare, K. C. Ho, and M. Skubic, "Multiple instance dictionary learning for beat-to-beat heart rate monitoring from ballistocardiograms," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2634–2648, Nov. 2018.
- [24] C. Jiao and A. Zare, "Functions of multiple instances for learning target signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4670–4686, Aug. 2015.
- [25] C. Jiao and A. Zare, "Multiple instance dictionary learning using functions of multiple instances," in *Proc. IEEE 23rd Int. Conf. Pattern Recognit.*, Dec. 2016, pp. 2688–2693.
- [26] C. Jiao, A. Zare, and R. G. McGarvey, (2017). "Multiple instance hybrid estimator for hyperspectral target characterization and sub-pixel target detection." [Online]. Available: <https://arxiv.org/abs/1710.11599>
- [27] A. Kirillov, B. Savchynskyy, D. Schlesinger, D. Vetrov, and C. Rother, "Inferring M-best diverse labelings in a single one," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1814–1822.
- [28] H. Kwon and N. M. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 178–194, Feb. 2006.
- [29] M. T. Law, Y. Yu, R. Urtasun, R. S. Zemel, and E. P. Xing, "Efficient multiple instance metric learning using weakly supervised data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 576–584.
- [30] S. Lee, S. Purushwalkam, M. Cogswell, V. Ranjan, D. Crandall, and D. Batra, "Stochastic multiple choice learning for training diverse deep ensembles," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2119–2127.
- [31] J. Ma and T. W. S. Chow, "Topic-based algorithm for multilabel learning with missing labels," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. doi: [10.1109/TNNLS.2018.2874434](https://doi.org/10.1109/TNNLS.2018.2874434).
- [32] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 1998, pp. 570–576.
- [33] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.
- [34] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. doi: [10.1109/TNNLS.2018.2874432](https://doi.org/10.1109/TNNLS.2018.2874432).
- [35] M. Qiao, L. Liu, J. Yu, C. Xu, and D. Tao, "Diversified dictionaries for multi-instance learning," *Pattern Recognit.*, vol. 64, pp. 407–416, Apr. 2017.
- [36] C. Zhang, J. C. Platt, and P. A. Viola, "Multiple instance boosting for object detection," in *Proc. IEEE Adv. Neural Inf. Process. Syst.*, vol. 18, 2005, pp. 1417–1424.
- [37] Z. Wang, L. Lan, and S. Vucetic, "Mixture model for multiple instance regression and applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2226–2237, Jun. 2012.
- [38] P. Williams, "Bayesian regularization and pruning using a Laplace prior," *Neural Comput.*, vol. 7, no. 1, pp. 117–143, 1995.
- [39] L. Xie, D. Tao, and H. Wei, "Early expression detection via online multi-instance learning with nonlinear extension," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. doi: [10.1109/TNNLS.2018.2869891](https://doi.org/10.1109/TNNLS.2018.2869891).
- [40] S. Yang, Z. Feng, M. Wang, and K. Zhang, "Self-paced learning-based probability subspace projection for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 630–635, Jun. 2018.
- [41] A. Zare, C. Jiao, and T. Glenn, "Discriminative multiple instance Hyperspectral target characterization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2342–2354, Oct. 2018.
- [42] Q. Zhang and S. Goldman, "EM-DD: An improved multiple-instance learning technique," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, 2001, pp. 1073–1080.
- [43] P. Zhong, Z. Gong, S. Li, and C.-B. Schönlieb, "Learning to diversify deep belief networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3516–3530, Jun. 2017.
- [44] P. Zhong, N. Peng, and R. Wang, "Learning to diversify patch-based priors for remote sensing image restoration," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5225–5245, Nov. 2015.
- [45] P. Zhong and R. Wang, "Learning conditional random fields for classification of hyperspectral images," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1890–1907, Jul. 2010.
- [46] P. Zhong and R. Wang, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1319–1334, Jul. 2014.

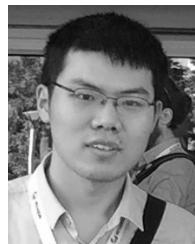


Ping Zhong (M'09–SM'18) received the M.S. degree in applied mathematics and the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2003 and 2008, respectively.

From 2015 to 2016, he was a Visiting Scholar with the Department of Applied Mathematics and Theory Physics, University of Cambridge, Cambridge, U.K. He is currently a Professor with the ATR National Laboratory, NUDT. He has authored

more than 30 peer-reviewed papers in international journals such as the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. His current research interests include computer vision, machine learning, and pattern recognition.

Dr. Zhong was a recipient of the National Excellent Doctoral Dissertation Award of China in 2011 and the New Century Excellent Talents in University of China in 2013. He is a Referee of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.



Zhiqiang Gong received the bachelor's degree in applied mathematics from Shanghai Jiaotong University, Shanghai, China, in 2013, and the master's degree in applied mathematics from the National University of Defense Technology, Changsha, China, in 2015, where he is currently pursuing the Ph.D. degree with the ATR National Laboratory.

His current research interests include computer vision, machine learning, and image analysis.



Jiaxin Shan received the bachelor's degree in communication engineering from the Harbin Institute of Technology, Harbin, China, in 2016. She is currently pursuing the master's degree with the ATR National Laboratory, National University of Defense Technology, Changsha, China.

Her current research interests include pattern recognition and machine learning.