





An Unmixing-Based Network for Underwater Target Detection From Hyperspectral Imagery

Jiahao Qi , Zhiqiang Gong , Wei Xue , Xingyue Liu, Aihuan Yao, and Ping Zhong , *Senior Member, IEEE*

Abstract—Detecting underwater targets from hyperspectral imagery makes a profound impact on marine exploration. Available methods mainly tackle this problem by modifying the land-based detection algorithms with classical bathymetric models, which usually fail to remove the interference of background and ignore the effect of depth information, leading to a poor detection performance. To achieve a more precise result, in this work we propose a novel network based on hyperspectral unmixing (HU) methodology and bathymetric models to detect the desired underwater targets. The proposed network, called underwater target detection network (UTD-Net), first develops a novel joint anomaly detector with classical HU methods to separate out target-water mixed pixels, which is devoted to eliminate the adverse influence of background. Then, we explore a bathymetric model-based autoencoder to unmix the target-water mixed pixels for acquiring the target-associated abundance values and maps. One dimension convolutional neural network is exploited to construct the encoder part of above autoencoder for the sake of addressing spectral variability problem. Moreover, considering the physical meaningless endmembers issue, a particular spectral constraint is imposed on the objective function as a training guidance. In this way, the autoencoder would be capable of generating specific endmembers and their corresponding abundance maps. Finally, according to the physical essence of abundance maps, we figure out the detection result by fusing the outcomes of autoencoder with weight coefficients determined by abundance values. Qualitative and quantitative illustrations demonstrate the effectiveness and efficiency of UTD-Net in comparison with the state-of-the-art underwater target detection methods.

Index Terms—Anomaly detection, bathymetric model-based (BMB) autoencoder, hyperspectral unmixing (HU), physical meaningless endmembers.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) possess affluent spectral information due to recording continuous spectra with high wavelength resolution, which are capable of depicting the

characteristics of different material surpassingly [1]–[4]. The reflectance values at different spectral bands of each pixel make up the reflectance spectrum that can always be regarded as a spectral feature vector. Since specific material generally conveys unique spectral feature vector named spectrum signature, HSIs are widely used in hyperspectral target detection tasks. In general, target detection can be interrupted as seeking out desired targets with prior information, which attracts remarkable attention in both civilian and military applications for several years [5]–[8].

Plenty of signature-based target detection algorithms have been proposed based on the assumption that target spectra are independent of the background. If the detection targets are underwater, aforementioned tasks will turn into underwater target detection. However, the target-background independence assumption is invalid owing to the interference derived from water environment under this context [17]. That is, existing signature-based detection algorithms might demonstrate unappealing performance in tackling underwater target detection problem. The main reason accounting for such a phenomenon is that any light reflected by target and captured by hyperspectral sensor has to pass through the water body [11]. Great attenuation will be added to the sensor-observed spectra by surrounding water column during above passing procedure, which is highly dependent on the inherent optical properties (IOPs) of water environment and the depth information of underwater target. Hence, identical target might have distinct sensor-observed spectra when the corresponding IOPs or depth information change.

Obviously, hyperspectral underwater detection remains a great challenge in remote sensing field, while a growing academic interest has been developed to occupy this challenge in recent years [9], [10], [12]–[15]. The major issue in hyperspectral underwater target detection is that the signature of desired target is usually uncertain with the impact of water environment, while most of existing detection methods ignore such important information. Consequently, prior knowledge of water environment such as IOPs should be introduced to restrict the variation of underwater target signature for addressing this issue. In other words, if the IOPs of water body and the depth information of target can be known beforehand, the required knowledge for signature-based detection methods will be available beforehand. Then, we could manage to find out the desired underwater targets with these signature-based methods. Moreover, the prevalent viewpoint to solve underwater target detection problem is predicting what the desired target would look like in a given underwater scenario before detection.

Manuscript received March 6, 2021; revised April 11, 2021; accepted May 11, 2021. Date of publication May 17, 2021; date of current version June 8, 2021. This work was supported in part by the Natural Science Foundation of China under Grant 61971428, Grant 61671456, Grant 61806004, and Grant 62001502, and in part by the China Postdoctoral Science Foundation under Grant 2020T130767. (*Corresponding author: Ping Zhong.*)

Jiahao Qi, Wei Xue, Xingyue Liu, Aihuan Yao, and Ping Zhong are with the National Key Laboratory of Science and Technology on Automatic Target Recognition, National University of Defense Technology, Changsha 410073, China, and also with the School of Computer Science and Technology, Anhui University of Technology, Maanshan 243032, China (e-mail: cswxue@ahut.edu.cn).

Zhiqiang Gong is with the National Innovation Institute of Defense Technology, Chinese Academy of Military Science, Beijing 100000, China (e-mail: gongzhiqiang13@nudt.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3080919

On account of these viewpoints, some related works have been proposed to address the underwater target detection problem. In [16], Jay *et al.* embedded bathymetric model into signature-based methods to detect underwater targets for both known and unknown water quality and designed a generalized likelihood ratio test (GLRT)-based bathymetric filter to retrieve the IOPs of water environment. It is notable that the GLRT always makes the hypothesis that unknown parameters should follow a multivariate normal distribution. However, in a practical underwater scene, the multivariate normal distribution might not be adequate to describe the correlation among different unknown parameters, which may have adverse impact on the generalization and effectiveness of underwater detector. Similarly, Gillis [17] proposed a general underwater target detection framework for hyperspectral imagery. This framework employs a spectral matching optimization algorithm (SMOA) to estimate water IOPs, and then constructs a target space model for depth information estimation. Unfortunately, only with sufficient prior knowledge about surrounding water can the SMOA achieve a promising performance. Furthermore, estimating the depth information of desired target might be computationally expensive, especially in depositing the large-scale datasets.

There is no doubt that remarkable achievements have been raised by the contribution of existing research works. But there still remains some crucial problems to be settled for further improving the detection performances in an underwater context. One major problem is that the detection performances of existing research works highly rely on the accuracies of IOPs retrieving results. Generally, it is necessary to collect enough prior information about the water environment beforehand for attaining an accurate IOPs retrieving result, which seems to be impossible in real-world applications. Besides, the prevalent standpoint adopted by most existing researches turns out to be removing the interference of water environment before detection [18]. However, this viewpoint does not take the spectral characteristics of underwater targets into consideration, which may undermine the accuracy of final detection result. Another universal problem is that the targets of an identical material are usually assumed to locate at the same depth in prior works for simplifying the underwater detection problem. However, this situation might be infrequent and the depth information usually has a great impact on the sensor-observed spectra of underwater targets, which is illustrated in Fig. 1. Therefore, a powerful underwater target detection algorithm must possess the capability of detecting the targets with different depth information in various water environments.

To detect the desired target in a given underwater context, a novel underwater target detection network (UTD-Net) has been proposed in this work. The proposed network consists of three significant modules: *Endmembers Separation Module*, *Underwater Target Separation Module*, and *Abundance Maps Fusion Module*. Primarily, to remove the negative effect of surrounding water environment, a joint anomaly detector is developed in the first module to separate the target-water mixed pixels from the background. Meanwhile, we also develop a cluster-based abundance average (CBAA) strategy to achieve a better separation result. Furthermore, based on the characteristics of target-water

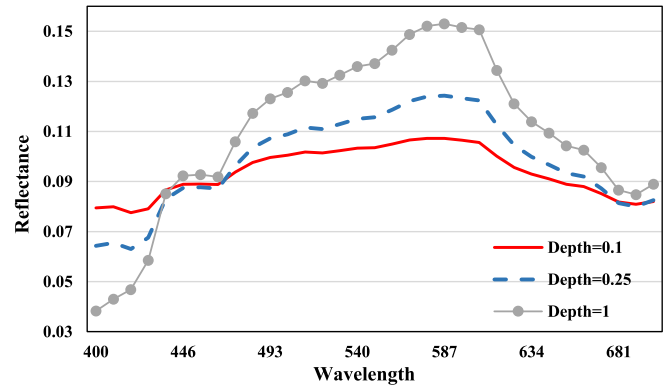


Fig. 1. Impact of depth information for underwater target spectrum.

mixed pixels, a special autoencoder is designed in the second module to unmix the output of joint anomaly detector for acquiring the suspicious target spectra and their corresponding abundance maps. In order to tackle the physical meaningless endmembers issue, we embed classical bathymetric model into the decoder part of the above autoencoder as an unmixing guidance. In addition, the embedding of bathymetric model can also make UTD-Net follow the same physical background as the existing underwater detection methods. Moreover, the 1-D convolutional neural network (1D CNN) is also exploited to establish the encoder part for the purpose of adapting spectral variability. Finally, we acquire the underwater detection result by fusing all the abundance maps of attained suspicious target endmembers based on their corresponding abundance values. The major contributions of this article are listed as follows.

- 1) A novel neural network named UTD-Net has been proposed to address underwater detection problem, which is based on deep learning and hyperspectral unmixing (HU) methodologies. To the best of our knowledge, it is the first time that these two methodologies are applied to detect desired underwater target in an underwater context.
- 2) To achieve a more robust target-water mixed pixels separation effect, we design a joint anomaly detection method with different classical HU methods. A CBAA strategy is proposed to integrate the anomaly detection results figured out by these HU methods. In this way, the joint anomaly detector can effectively eliminate the influence of background and attain a promising detection result.
- 3) In order to address the physical meaningless endmembers issue, we put forward a novel constraint for the objective function of bathymetric model-based (BMB) autoencoder. This specific constraint acts as an guidance to restrict the range of unmixing results. Meanwhile, the decoder part of autoencoder is constructed with 1D CNN which devotes to tackling the spectral variability problem derived from the environment factors.

The remainder of this article is organized as follows. Section II briefly reviews the essential knowledge used in our research. In Section III, we introduce all the details about the proposed UTD-Net. Section IV shows the performance of our

proposed method on synthetic sets. Section V, concludes this article.

II. PRELIMINARIES

In this section, we briefly introduce the bathymetric model developed in hyperspectral oceanography, which is essential to our research work. To further demonstrate the proposed method, hyperspectral mixing modeling is also mentioned in this section.

A. General Bathymetric Model

Generally speaking, the bathymetric model is devoted to formulating the mechanism of sensor-observed underwater spectra [19]–[21]. It is self-evident that the sensor-observed spectra are generated by the sunlight which has succeeded in passing through a water body. Intuitively, the process of generating such sensor-observed spectra can be depicted as follows: The sunlight first enters into the water body, then might be reflected by underwater targets or water columns, and subsequently be captured by hyperspectral sensors at the end. Furthermore, a mass of attenuation will be imposed on sunlight during such transmission process. It is worth mentioning that the spectra captured by sensors compose of the reflectance spectra from underwater targets and the reflectance spectra from water body. On the basis of [23], the essence of sensor-observed underwater spectra is a linear combination of two different kinds of reflectance spectra with the weight coefficients determined by water attenuation. Consequently, general bathymetric model can be defined as

$$r(\lambda) = r_{\infty}(\lambda) \left(1 - e^{-2k(\lambda)H}\right) + \frac{r_B(\lambda)}{\pi} e^{-2k(\lambda)H} \quad (1)$$

where $r(\lambda)$ denotes the sensor-observed underwater spectrum, $r_{\infty}(\lambda)$ and $r_B(\lambda)$ refer to the reflectance spectrum from the water body and the reflectance spectrum from underwater target severally. In addition, H is the depth information of underwater target.

With (1), we can comprehensively understand the constituents of a sensor-observed underwater spectrum and gather more useful information for underwater target detection. The former term in (1) corresponds to the contribution of water body, whereas the later one represents the contribution derived from underwater target. And it is noticeable that the weight coefficients of these two term possess sum-to-one and nonnegative characteristics.

B. Hyperspectral Linear Mixing Models

Due to the limitation of spatial resolution, many pixels consist of several materials, which brings strong interference to HSI understanding [24], [27]. HU is one crucial technology to address this problem that aims at separating the mixed pixels into a set of spectral signatures $\{e^m\}_{m=1}^M$, termed as endmembers. Each endmember possesses a corresponding abundance a_m representing the areal abundance fraction of this endmember. In this way, we can depict a mixing model as follows:

$$\mathbf{x} = f(\mathbf{E}, \mathbf{a}) \quad (2)$$

where \mathbf{E} represents the matrix form of endmembers set $\mathbf{E} = (e^1, \dots, e^M)$ and $\mathbf{a} = (a^1, \dots, a^M)^T$ refers to the abundances set. The function f , describing how to obtain a mixed pixel \mathbf{x} with \mathbf{E} and \mathbf{a} , also decides the type of mixing model to be linear or nonlinear.

Linear mixing model (LMM) and nonlinear mixing model (NLMM) are the major types of mixing models [28]. Definitely, the NLMM is closer to the real scenarios. However, taking care about computation complexity, most of the literatures tend to address hyperspectral mixing problems with LMM [29]. The LMM holds that sensor-observed spectrum is a weighted average of the light scattered from various materials. Despite its simplicity, it represents an acceptable approximation in most real scenarios which can also dramatically simplify the complexity of unmixing problems. Apart from the linear scattering mechanism for endmembers, LMM also imposes two constraints on abundances: 1) Abundance nonnegativity constraint (ANC), restricting every abundance to be nonnegative; and 2) abundance sum-to-one constraint (ASC), commanding that the fractional abundances of the endmembers must be summed to one. Putting everything together, this prevalent and physical interpretable mixing model can be formulated as follows:

$$\begin{aligned} \mathbf{x} &= \sum_{m=1}^M a^m e^m + \mathbf{n} \\ &= \mathbf{E}\mathbf{a} + \mathbf{n} \end{aligned} \quad (3)$$

where \mathbf{n} represents the noise vector that has a great impact on unmixing results. From (3), we can further understand the constituents of mixed pixels. Moreover, the noise vector \mathbf{n} refers to additional uncorrelated Gaussian noise, which can be eliminated with traditional denoising methods.

III. UTD-NET: THE UNDERWATER TARGET DETECTION NETWORK

In this section, we elaborate on the proposed method UTD-Net for underwater target detection. UTD-Net consists of three parts which is illustrated in Fig. 2. The first part refers to *endmembers extraction module* which conducts a joint anomaly detector based on classical HU methods. And this detector will be employed to separate the water-target mixed pixels from the background for removing the influence of water environment. Then, a BMB autoencoder is developed in the second part *underwater target separation module* of UTD-Net. With this specific autoencoder, we are capable of deriving desired target spectra from mixed pixels regardless of the affection from water IOPs and target depth information. After that, an abundance maps fusion method has been proposed in the last part, *abundance maps fusion module*, which is designed for figuring out the final underwater detection result. This method tends to endow clustered abundance maps with different weights based on their corresponding abundance values. Finally, fusing all the derived clustered abundance maps with their weights, a promising underwater detection result is available in any underwater scenario.

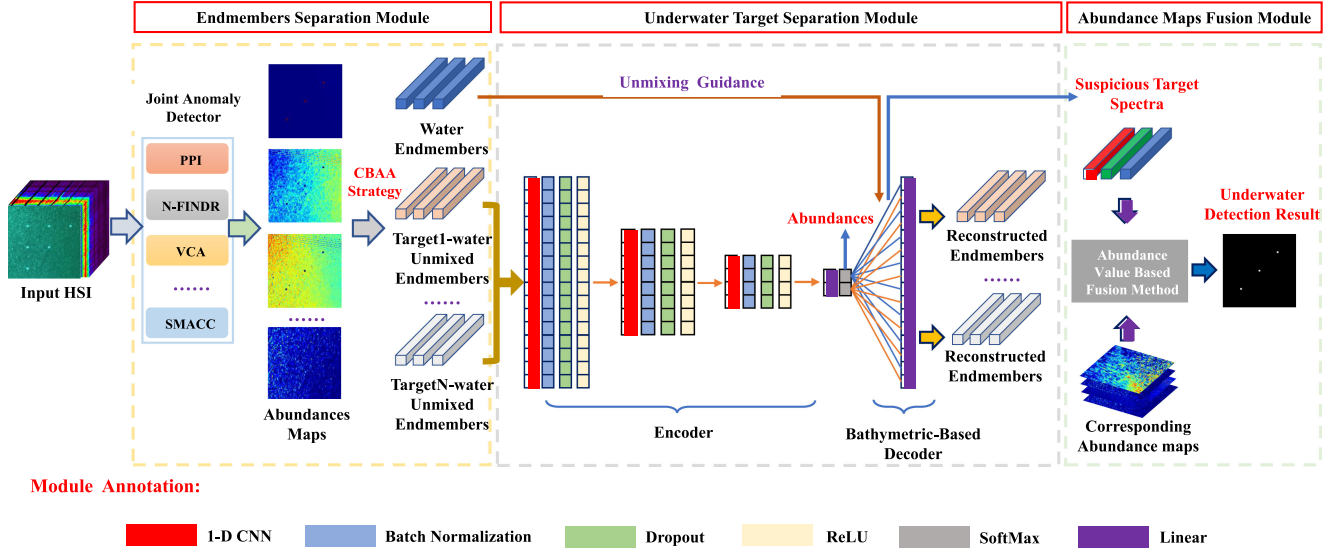


Fig. 2. Diagram of the proposed method.

A. Endmembers Separation Module

According to related research works [25], [26], targets pixels usually occupy a tiny percent of the total HSI in hyperspectral target detection task. Following this viewpoint, target-water mixed pixels in the underwater HSI own the sparse characteristic as well. Meanwhile, if the underwater targets are detectable, the target-water mixed pixels will exhibit distinctive features compared with surrounding water body background. Consequently, based on [30], it is reasonable to consider the target-water mixed pixels as outliers in underwater scenario. However, from (1) we can know that, unlike the general ones existing in land-based anomaly detection problems, these outliers are highly dependent with the background pixels. Hence, traditional hyperspectral anomaly detection methods, detecting outliers based on the statistical characteristics of dataset, are invalid while employing them to detect target-water mixed pixels from the given hyperspectral water imagery.

HU methods, merely taking the spectral discrepancy as criterion to transform input HSI into different endmembers, are capable of finishing above outliers detection mission. Moreover, HU methods can achieve a satisfying performance even if the target-water mixed pixels are not so distinguishable with their background pixels. Consequently, we design an anomaly detector based on HU methods in this work.

Glancing over the architecture of UTD-Net, *endmembers extraction module* is the most significant part which determines the final detection result directly. To further find out the target-water mixed pixels, a joint anomaly detection method integrating various HU methods has been proposed in this module. The flowchart of this novel detector is illustrated in Fig. 3. Let $\mathbf{X} \in \mathbb{R}^{L \times W \times B}$ be the matrix form of input HSI which possesses B bands with $L \times W$ spatial solution. At the very beginning, the maximum noise fraction (MNF) algorithm is employed to reduce the dimension of input HSI while eliminating the additional uncorrelated Gaussian noise which is mentioned in (3). Then,

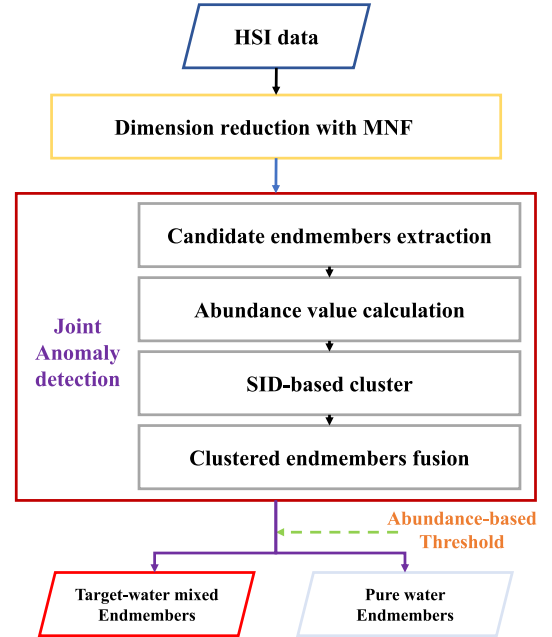


Fig. 3. Flowchart of joint anomaly detector.

for the preprocessed HSI $\hat{\mathbf{X}} \in \mathbb{R}^{L \times W \times b}$, anomaly detection is performed by different typical HU methods

$$(\mathbf{E}^k, \mathbf{A}^k) = \mathbf{f}^k(\hat{\mathbf{X}}) \quad (4)$$

where \mathbf{f}^k refers to the k th method in HU methods set $\{\mathbf{f}^1, \dots, \mathbf{f}^K\}$. $\mathbf{E}^k = (e^{k,1}, \dots, e^{k,p})^T \in \mathbb{R}^{p \times B}$ represents endmember matrix containing p endmembers. $\mathbf{A}^k = (a^{k,1}, \dots, a^{k,p})^T \in \mathbb{R}^{L \times W \times p}$ is the corresponding abundance matrix. The symbol k indicates that this result is generated by the k th HU method. In abundance matrix, $a^{k,n} \in \mathbb{R}^{L \times W}$ represents the abundance map of the n th endmember, where $n = 1, 2, \dots, p$. And the abundance value of the n th endmember

generated by the k th HU method can be calculated as

$$a^{k,n} = \frac{\sum_{i=1}^L \sum_{j=1}^W a_{ij}^{k,n}}{(L \times W)} \quad (5)$$

where $a_{ij}^{k,n}$ is the entry in the i th row and j th column of the abundance map matrix $\mathbf{a}^{k,n}$. L and W refer to the length and width of the abundance map matrix, respectively. As for the physical essence of $a^{k,n}$, it denotes the percentage occupied by its corresponding endmember among the whole HSI. Consequently, for the outcome of each HU method, we can rewrite it as the endmember-abundance value pair form

$$(\mathbf{E}^k, \mathbf{A}^k) \Rightarrow \{(e^{k,1}, a^{k,1}), \dots, (e^{k,p}, a^{k,p})\}. \quad (6)$$

Since there is no prior knowledge about the amount of endmembers, in this article we will utilize each HU method to attain p endmembers ($p = nq$, where q represents the number of underwater target that need to be detected). And a collection of endmembers $\mathbf{E}_{\text{all}} = \{e^m\}_{m=1}^{K \times p}$ is available after handling the preprocessed HSI $\hat{\mathbf{X}}$ with K different HU methods. It is worth mentioning that different HU methods will get disparate results even if the input HSI is identical. Meanwhile, if the water environment condition changes a lot among the input HSI, water background pixels might be considered as different endmembers. To get around these issues for achieving a more robust and effective anomaly detection performance, we fuse these different HU results to attain a blending decision.

Considering the characteristics of HU methods, a CBAA strategy is proposed as follows. All the derived endmembers and their corresponding abundances can be grouped into p clusters $\{C^i\}_{i=1}^p$ based on learning vector quantization (LVQ) cluster method [31] and spectral information divergence (SID) metric [32]. Let C_c^i and C_c^j be the cluster centers of C^i and C^j . After clustering by SID-based LVQ method, any endmember e that belongs to C^i will form an endmember collection

$$\{e \in \mathbf{E}_{\text{all}} \mid \text{SID}(e, C_c^i) \leq \text{SID}(e, C_c^j), j \neq i\} \quad (7)$$

where SID is an excellent spectral distance metric that devotes to comparing the relative discriminatory power between two spectra. This specific distance metric can be formulated as follows:

$$\text{SID}(\mathbf{x}, \mathbf{y}) = D(\mathbf{x} \parallel \mathbf{y}) + D(\mathbf{y} \parallel \mathbf{x}) \quad (8)$$

where \mathbf{x} and \mathbf{y} are the sequence forms of two pixels required to capture the spectral correlation. $D(\mathbf{x} \parallel \mathbf{y})$ represents the Kullback–Leibler information measure between the input pixels.

The final operation in CBAA strategy is to derive the fusion result from different clusters. As demonstrated in (6), each endmember has its corresponding abundance value, which indicates the importance of this endmember for the whole HSI. Therefore, we denote the linear combination of all the endmember elements in one cluster as its endmember fusion result with weight coefficients depended on abundance values. The final fusion results for all the clusters are demonstrated as follows:

$$\hat{e}^i = \frac{\sum_{n=1}^N e_n^i \cdot a_n^i}{\sum_{n=1}^N a_n^i}, \quad i = 1, 2, \dots, p \quad (9)$$

where \hat{e}^i denotes the fusion result of the i th endmember cluster containing N endmembers. e_n^i and a_n^i represent the n th endmember and its corresponding abundance value in the i th endmember cluster. p refers to the amount of endmember clusters. Meanwhile, the associated map values of fusion endmembers are generated with the following rule:

$$\hat{\mathbf{A}}^i = \frac{\sum_{n=1}^N \mathbf{A}_n^i}{K \times p}, \quad i = 1, 2, \dots, p \quad (10)$$

where \mathbf{A}_n^i represents the abundance map of the n th endmember in the i th endmember cluster and Kp is the total quantity of the endmembers produced by the K different HU methods. With (5), (9), and (10), we will attain p novel {endmember, abundance value, abundance map} pairs

$$\{(\hat{e}^1, \hat{a}^1, \hat{\mathbf{A}}^1), \dots, (\hat{e}^p, \hat{a}^p, \hat{\mathbf{A}}^p)\}. \quad (11)$$

Note that, there are only q targets that need to be detected, and hence we should pick out some particular endmembers from the fusion result. Obviously, the water background endmember occupies most percent of the total HSI to acquire the largest abundance value. Therefore, we can get the pure water endmembers by picking out the endmembers whose abundance value is the largest one. Then, the reminder of endmembers will be regarded as candidate water-target mixed endmembers which will be sent to subsequent modules. A pseudocode of the proposed joint anomaly detection method is given in Algorithm 1.

B. Underwater Target Separation Module

With the effort of joint anomaly detection, pure water endmember and candidate water-target mixed endmembers are available. Then, we have to sort out all the water-target mixed endmembers from candidates and employ them to figure out the final underwater detection result. Unfortunately, due to the effect of surrounding water environment, the same material locating in different positions will be associated with distinct endmembers. It is noticeable that each endmember of underwater HSI must contain the pure water spectrum according to the analysis provided in Section II. Meanwhile, as it is demonstrated in (1), target-mixed endmember turns out to be a linear combination of pure water spectrum and target spectrum with coefficients matching sum-to-one and nonnegative constraints. Therefore, if the candidate target-water mixed endmember can be unmixed into water spectrum and another unidentified spectrum, we can utilize the signature of this unidentified spectrum to decide whether its corresponding endmember belongs to water-target mixed endmembers.

Specifically speaking, the HU methodology can be used once again to tackle the aforementioned problems or challenges. However, traditional HU methods may not be satisfied in this situation since their HU processes are uncontrollable. More precisely, they might product physically meaningless endmembers due to lacking of effective guidance about real endmembers in the blind HU. Autoencoder is an excellent deep neural network structure widely adopted in HU research work whose HU process can be controllable with specific training guidance.

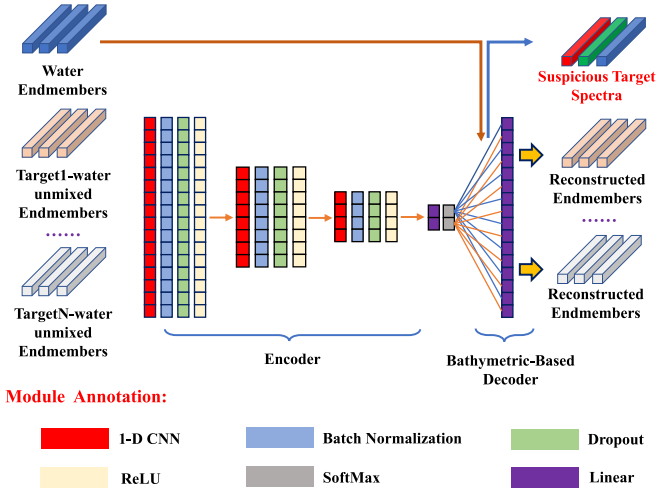


Fig. 4. Diagram of the BMB autoencoder.

Following this viewpoint, to address the physically meaningless endmembers issue, we develop a BMB autoencoder illustrated in Fig. 4 and the proposed autoencoder composes of two principle parts named encoder and decoder.

In encoder part, the input mixed endmember e will be compressed into a low-dimension vector v which represents the abundance vector. The physical meaning of compression procedure can be interrupted as estimating the abundance value for each endmember according to its unique spectral features. Considering the effect of spectral variability derived from illumination, atmospheric conditions, and other environmental factors, we conduct backbone network with 1D CNN as spectrum feature extractor. The 1D CNN which has been exploited to attain middle-level, locally invariant, and discriminative features from the input spectrum can efficiently eliminate the interference of spectral variability. Let $e \in \mathbb{R}^b$ be input endmember with b elements, the output of the t layer in backbone network is defined as

$$\mathbf{y}^{(t)} = \begin{cases} f(\mathbf{w}^{(t)} * \mathbf{e} + \mathbf{b}^{(t)}), & t = 1 \\ f(\mathbf{w}^{(t)} * \mathbf{y}^{(t-1)} + \mathbf{b}^{(t)}), & t \geq 2 \end{cases} \quad (12)$$

where $*$ refers to the convolution math operation and f represents the nonlinear activation function contributing to impose nonlinearity on encoder network. In this article, to make abundance vector obey ANC, ReLU is utilized as the activation function. Furthermore, the batch normalization trick is also introduced to improve robustness for gradient vanishing and speediness of convergence before activation operation. After 1D CNN-based backbone network, a fully connected layer is established to restrict the dimension of output vector v . According to (1), we can know that there exists only two endmembers in target-water mixed endmembers. Therefore, the dimension of v will be set to 2 for matching the endmembers amount of target-water mixed endmembers. Moreover, the softmax function is exploited as the activation function for fully connected layer to

Algorithm 1: Joint Anomaly Detection Method.

Input: HSI X , HU methods set $\{\mathbf{f}^i\}_{i=1}^K$

- 1: **for** $i = 1, 2, \dots, K$ **do**
- 2: Compute (E^k, A^k) according to (4);
- 3: **for** $j = 1, 2, \dots, p$ **do**
- 4: Obtain the abundance value for $e^{i,j}$ via (5);
- 5: **end for**
- 6: **end for**
- 7: Attain p endmember clusters $\{C^i\}_{i=1}^p$ based on SID-based LVQ method;
- 8: **for** $n = 1, 2, \dots, p$ **do**
- 9: Get the novel endmember \hat{e}_n with (9);
- 10: Update corresponding abundance map \hat{A}^n by (10);
- 11: Calculate associated abundance value \hat{a}_n via (5);
- 12: **end for**
- 13: Acquire candidate target-water mixed endmembers and pure water endmember according to abundance value based threshold.

Output: Target-water mixed pixels and their corresponding abundance maps

satisfy ASC and the output vector v can be calculated as follows:

$$\mathbf{v}_i = \frac{e^{\mathbf{y}_i^{(\text{output})}}}{\sum_{k=1}^2 e^{\mathbf{y}_k^{(\text{output})}}} \quad (13)$$

where $\mathbf{y}_k^{(\text{output})}$ denotes the k th element in the outperformance of fully connected layer which is also associated with the k th abundance value.

With respect to the decoder, it aims at recovering abundance vector v to reconstruct the input mixed endmember e . In this work, the decoder consists of merely one fully connected layer while the relationship between reconstructed endmember \tilde{e} and abundance vector v is depicted as follows:

$$\tilde{e} = f(v) = f(\mathbf{w}^d \mathbf{v} + \mathbf{b}^d) \quad (14)$$

where \mathbf{w}^d and \mathbf{b}^d are the weight matrix and bias term between v and reconstructed endmember \tilde{e} separately. Similarly, f represents activation function and the linear function will be used in the decoder part. Therefore, we can rewrite (14) into a more concrete form

$$\tilde{e} = \mathbf{w}^d \mathbf{v} + \mathbf{b}^d. \quad (15)$$

It is obvious that this formula matches the linear mixing model exhibited in (3), simultaneously v also guarantees the ANC and the ASC. Putting it in another way, once the autoencoder has been properly trained, the unmixing process can be performed by

$$\text{Endmembers Extraction : } \hat{E} \simeq \mathbf{w}^d \triangleq \{\mathbf{w}_1^d, \mathbf{w}_2^d\} \quad (16)$$

$$\text{Abundance Estimation : } \hat{a} \simeq \mathbf{v} \triangleq \{a_1, a_2\}. \quad (17)$$

However, such a process only unmixes the input pixels without taking consideration about the physical essence of the unmixing result. As mentioned, an appropriate guidance is required for blind unmixing which is devoted to insuring one of the

extracted endmembers to be the water spectrum. To achieve this goal, we add a BMB constraint into objective function. The BMB constraint first picks out a proper weight vector contained in weight matrix \mathbf{w}^d that owns smaller spectral distance with reference water spectrum. Then, it tries to decrease the spectral distance between the selected weight vector and reference water spectrum. The proposed constraint can be formulated as

$$f(\mathbf{w}^d) = \|\tilde{\mathbf{w}}^d - \mathbf{r}_\infty\|_2 \quad (18)$$

with

$$\tilde{\mathbf{w}}^d = \arg \max_{\mathbf{w} \in \mathbf{w}^d} \|\mathbf{w} - \mathbf{r}_\infty\|_2 \quad (19)$$

where $\|\cdot\|_2$ denotes the l_2 norm and \mathbf{r}_∞ is the reference water spectrum derived from joint anomaly detector. After that, we can tackle the simultaneous estimation of $\hat{\mathbf{E}}$ and $\hat{\mathbf{a}}$ with BMB guidance by finding out an optimal solution for the following optimization problem:

$$(\mathbf{w}^d, \mathbf{v}) = \arg \min_{\mathbf{w}^d, \mathbf{v}} \frac{1}{2} \|\hat{\mathbf{e}} - \mathbf{w}^d \mathbf{v}\|^2 + \lambda f(\mathbf{w}^d) \quad (20)$$

where $\hat{\mathbf{e}}$ refers to the input mixed endmember. The former term in (20) represents the reconstruction error and the later term denotes the BMB constraint. With this BMB autoencoder, we can transform any candidate target-water mixed endmember into a water spectrum $\hat{\mathbf{r}}_\infty$ and a suspicious target spectrum \mathbf{r} . Analogously, the abundance value a_s of suspicious target spectrum can also be selected based on the spectral distance with reference water spectrum from abundance estimation result $\hat{\mathbf{a}}$. In conclusion, with the contribution of BMB autoencoder, the input pairs mentioned in (11) will be transformed into $p - 1$ novel suspicious target spectrum and abundance value pairs $\{(a_s^1, \mathbf{r}^1), \dots, (a_s^p, \mathbf{r}^p)\}$.

C. Abundance Maps Fusion Module

As exhibited in (11), each endmember $\hat{\mathbf{e}}^i$ has its corresponding abundance map $\hat{\mathbf{A}}^i$. Moreover, the suspicious target spectra generated in the last subsection are derived from the endmembers in (11). Taking such relationship into consideration, we can conclude that $\hat{\mathbf{e}}^i$ and \mathbf{r}^i should share the identical abundance map $\hat{\mathbf{A}}^i$. Furthermore, considering the physical essence of the abundance value a_s , it can be interrupted as the proportions of target spectrum in a mixed endmembers. In some way, a_s represents the probability of being a target spectrum for its associated mixed endmembers. Meanwhile, the depth information of underwater targets is also connected with a_s . To put the (1) and (16) together, a_s turns out to be the coefficient of the second term in bathymetric model

$$a_s = e^{-2k(\lambda)H}. \quad (21)$$

When the depth information H becomes larger, the target will be more undetectable while the abundance value a_s becoming smaller, and vice versa. Consequently, the effect of depth information is reflected on the abundance value a_s . Moreover, the abundance maps demonstrate the areal abundance fraction of their corresponding endmembers which are approximate to the detection result maps. Consequently, we can achieve the final

underwater detection result by fusing all the abundance maps of suspicious target spectra with weight coefficients determined by the abundance values

$$\mathbf{D} = \sum_{i=1}^{p-1} \frac{\hat{\mathbf{A}}^i \cdot a_s^i}{\sum_{j=1}^{p-1} a_s^j}. \quad (22)$$

If a mixed endmember $\hat{\mathbf{e}}$ is similar to the water background pixels, it will be more impossible to contain the underwater target and its related abundance value a_s will be very small. According to (22), the abundance map of this endmember will contribute less to the final detection results. Besides, the final detection result figured out in a fusion form can reduce the false alarm rate while increasing the detection performance.

IV. EXPERIMENTS

In this section, we perform adequate experiments on synthetic sets associated with different underwater scenarios to exhibit the validity of the UTD-Net. At the very beginning, we will briefly describe the indispensable information about the employed datasets. Second, the evaluation criteria and parameter settings about the experiments are listed in the subsection *Experiment Details*. After that, to demonstrate the innovativeness of our proposed method, we design some specialized tests to validate the effectiveness of HU methods, joint anomaly detector and BMB constraint. Moreover, the fourth subsection is mainly about the underwater target detection experiment results and their corresponding discussion. Finally, the remainder of this section is devoted to conducting some specific experiments for exhibiting a more comprehensive analysis for the UTD-Net.

A. Hyperspectral DataSets

In an ideal situation, we could accomplish all the experiments with the real datasets. However, owing to the novelty of hyperspectral underwater research, none of public hyperspectral datasets containing underwater targets. To get around such dilemma, we adopt the strategy mentioned in [17] to generate the hyperspectral underwater targets detection datasets as realistic as possible. The data generating progress can be divided into several procedures as follows.

- 1) Select real-world hyperspectral water image as background and figure out their IOPs.
- 2) Choose appropriate target spectra from the USGS spectral library [22] and produce the underwater target spectra with the bathymetric models in (1).
- 3) Add a white Gaussian noise onto the generated underwater spectra to create the intraclass variability.
- 4) Embed the noisy underwater spectra into the background.

In order to comprehensively show the outperformance of UTD-Net, we collect four different types of hyperspectral water images as backgrounds. More detailed information about testing datasets is shown below.

1) *Synthetic Turbid Water Dataset*: The first testing dataset is created by a synthetic turbid water whose IOP parameters are measured in [33] and the sheet metal material spectrum. This dataset possesses 100×100 spatial resolution and its wavelength

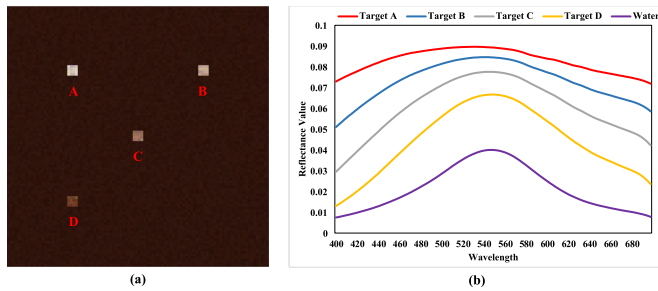


Fig. 5. Detailed information of synthetic water dataset.

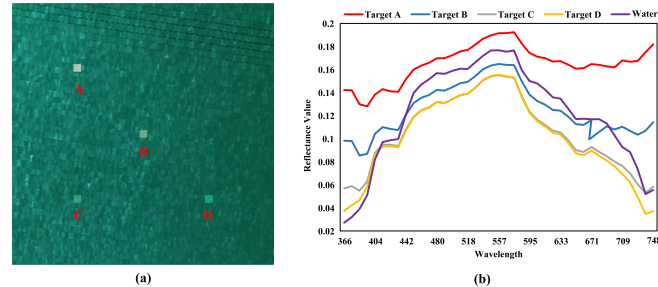


Fig. 6. Detailed information of sea water dataset.

ranging from 400 to 700 nm with 150 bands. Furthermore, the noisy underwater targets are placed at various depths 0.1, 0.25, 0.5, and 1 m and the corresponding target size is 5×5 . The concrete conditions of the synthetic turbid water dataset are illustrated in Fig. 5.

2) *Sea Water Dataset*: The second dataset is derived from a hyperspectral sea water image which was collected by the airborne visible infrared imaging spectrometer (AVIRIS). The collection position is a gullet locating in Galveston Bay, Texas, United States of America. This HSI covers 366 to 2495 nm at 9.5-nm spectral resolution and we segment out a chip with 342×342 pixels for experiment. In terms of the target spectra, they are the reflectance spectra of alunite material. The depth information of underwater target spectra is 0.1, 1, 2.5, and 5 m and the target size is 7×7 . We demonstrate the specific information about this dataset in Fig. 6.

3) *Lake Water Dataset*: The background of third dataset is a scene of Dongting Lake in Yueyang City, Hunan Province, China. The HSI is captured by Gaofen-5 satellite with advanced hyperspectral imagery (AHSI) in 2020 which has 330 spectral bands covering from 400 to 2500 nm. Similarly, a 60×60 image chip is selected to conduct the underwater target detection dataset and the target is particle board material. Moreover, the target depths are in the range of 0.1 to 4 m with an alterable step size and the target size is set as 3×3 . Fig. 7 has detailedly displayed the associated information about this dataset.

4) *River Water Dataset*: The last dataset is composed of nylon webbing material and hyperspectral river water image. The hyperspectral river water image was captured by Nano-Hyperspec sensor and the experiment location is Nangang River, Guangzhou City, Guangdong Province, China. Compared with previous hyperspectral water images, this one covers a narrow spectral range, whose wavelengths range from 400 to 1000 nm at

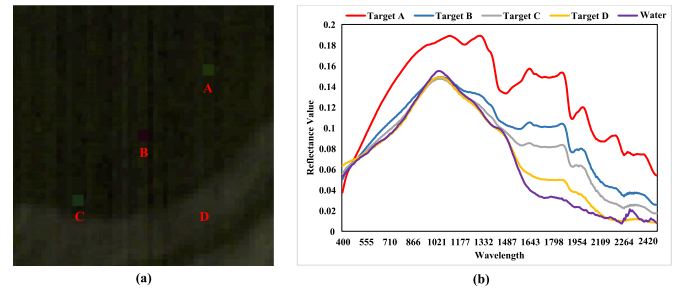


Fig. 7. Detailed information of lake water dataset.

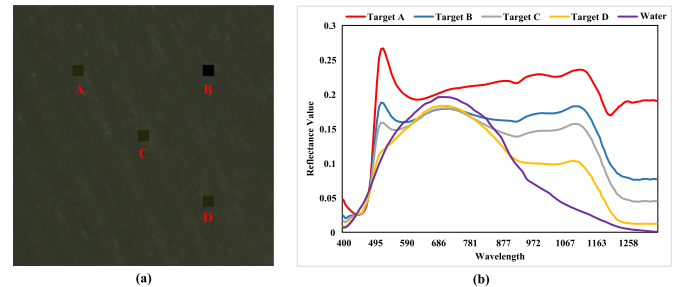


Fig. 8. Detailed information of river water dataset.

TABLE I
SUMMARY OF IMPORTANT INFORMATION FOR ALL THE
EXPERIMENTAL DATASETS

Dataset	Spatial solution	Size (shape)	Material	Depth information
Synthetic water	100×100	5×5	Sheet metal	0.1m, 1m, 2m, 3m
Sea water	384×384	7×7	Alunite	0.5m, 1m, 2.5m, 5m
Lake water	60×60	3×3	Particle board	0.1m, 1m, 2m, 5m
River water	180×180	5×5	Nylon webbing	1m, 3m, 5m, 7m

2.22-nm spectral resolution. In addition, considering the turbidity of the lake, we only set the targets at some shallow depths 0.1, 0.25, 0.5, and 0.75 m and the associated target sizes as 5×5 . A 180×180 pixel chip has been sampled in our experiment which will be demonstrated in Fig. 8.

In summary, the vital information about all the datasets is listed in Table I.

B. Experiment Details

To better introduce the experiment results for confirming the superiority of our research work, the essential experiment information will be presented in this section. First of all, we will shortly describe the evaluation criteria exploited to measure the performance of different testing methods. After that, the experiment settings are also listed at the end of this section.

1) *Evaluation Criteria*: In order to evaluate the performance of underwater target detection task, we first use the receiver operating characteristic (ROC) curve as a criterion. ROC curve has been regarded as one of the most widely used evolution metrics in vision detection missions which tends to depict the relationship between target detection probability P_d and the false alarm rate P_f [34]. These two important parameters can be

TABLE II
DETAILED INFORMATION ABOUT TRAINING, VALIDATION AND TESTING SETTINGS

Data set	The numbers of pixels		
	Training data set	Validation data set	Testing data set
Synthetic water	7000	1000	2000
Sea water	81874	11696	23392
River water	2520	360	720
Lake water	22680	3240	6480
Overall	114074	16296	32592

TABLE III
AUC VALUES FOR EFFECTIVENESS EVALUATION OF DIFFERENT ANOMALY DETECTORS

Data set	AUC values of different anomaly detection methods			
	RX [41]	GMRX [42]	CSD [43]	Joint Anomaly Detector
Synthetic water	0.7418	0.7725	0.7541	0.8637
Sea water	0.6947	0.7283	0.7302	0.8291
River water	0.7514	0.7998	0.7786	0.8949
Lake water	0.6582	0.7342	0.7219	0.8483
Average	0.7115	0.7587	0.7462	0.8590

The bold entries represent the best performance in each row.

defined as follows:

$$P_d = \frac{N_0}{N_t} \quad (23)$$

$$p_f = \frac{N_1}{N} \quad (24)$$

where N_0 and N_1 denote the amount of detected target pixels and the number of background pixels mistaken as targets under a specific threshold. N_t refers to be the total amount of desired target pixels while N represents the gross number of all the pixels in the testing HSI. We are capable of acquiring the comparison results about the detection performances for different methods in vision with ROC curves. In addition, to achieve the quantitative analysis about detection results, the area under ROC curve (AUC) value has also been calculated in our experiments.

2) *Experiment Settings*: For all the datasets, the standard deviation of white Gaussian noise is set as 0.05 to create a certain intraclass variability. At the same time, some particular models (such as ATCOR model [35]) are implemented on these datasets to eliminate the influence of environment factors. For establishing a more stable joint anomaly detector, we utilize the pure pixel index [36], N-FINDR [37] and vertex component analysis [38] methods as the fundamental algorithms. Moreover, so as to preferably measure the effectiveness of the proposed methods, UTD-Net will make the comparison with the following algorithms: 1) UTDF [17]; 2) GBF [16]; 3) CEM [39]; and 4) MF [40]. For all the testing methods, they are fed with the identical land-based target spectral signatures as the prior detection information.

Besides, the training, validation and testing settings are listed in Table II. It is noticeable that in all the datasets, 70% of pixels

TABLE IV
SPECTRAL DISTANCE INFORMATION FOR EFFECTIVENESS EVALUATION OF BMB CONSTRAINT

Data set	$D\langle r_t(\lambda), \hat{r}_t(\lambda) \rangle$		$D\langle r_\infty(\lambda), \hat{r}_\infty(\lambda) \rangle$	
	UTD-Net(Y)	UTD-Net(N)	UTD-Net(Y)	UTD-Net(N)
Synthetic water	0.01512	0.178	0.0289	0.2228
Sea water	0.08746	0.1871	0.0125	0.1498
River water	0.01511	0.2421	0.0459	0.2897
Lake water	0.04254	0.1932	0.0951	0.1393
Average	0.04006	0.2001	0.0456	0.2004

The bold entries represent the best performance in each row.

TABLE V
AUC VALUES OF UNDERWATER TARGET DETECTION RESULTS FOR COMPARED METHODS ON ALL THE DATASETS

Data set	Area Under ROC curves				
	CEM	MF	GBF	UTDF	UTD-Net
Synthetic water	0.5473	0.5491	0.6685	0.71294	0.95848
Sea water	0.5561	0.5501	0.7469	0.7187	0.92701
River water	0.6061	0.5651	0.81613	0.84134	0.99995
Lake water	0.52136	0.5602	0.72946	0.84391	0.93771
Average	0.5577	0.5561	0.7403	0.7792	0.9558

The bold entries represent the best performance in each row.

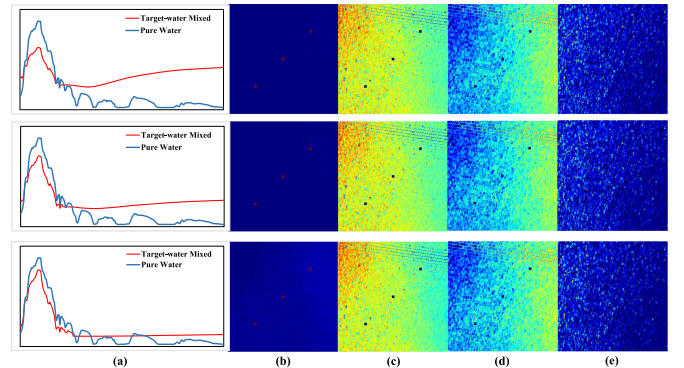


Fig. 9. Unmixing results of identical underwater target in different depth. (a) Spectral information. (b)–(e) Abundance maps.

are employed as the training samples, 10% of pixels are sampled to establish validation set, and the rest pixels are regarded as testing samples. Note that, all the samples are selected with the stratified random sampling strategy from input HSI.

Finally, related experiments are performed with the support of an Intel(R) Core (TM) i9-10920X CPU machine and 64 GB of RAM on Windows 10 operating system.

C. Component Analysis

In this section, we will accomplish some specialized tests to prove the validity of viewpoints mentioned beforehand for the UTD-Net.

1) *Effectiveness Evaluation of HU Method*: In the Section III-A, we have stated that HU method can separate the target-water mixed pixels. To validate this viewpoint, we have conducted

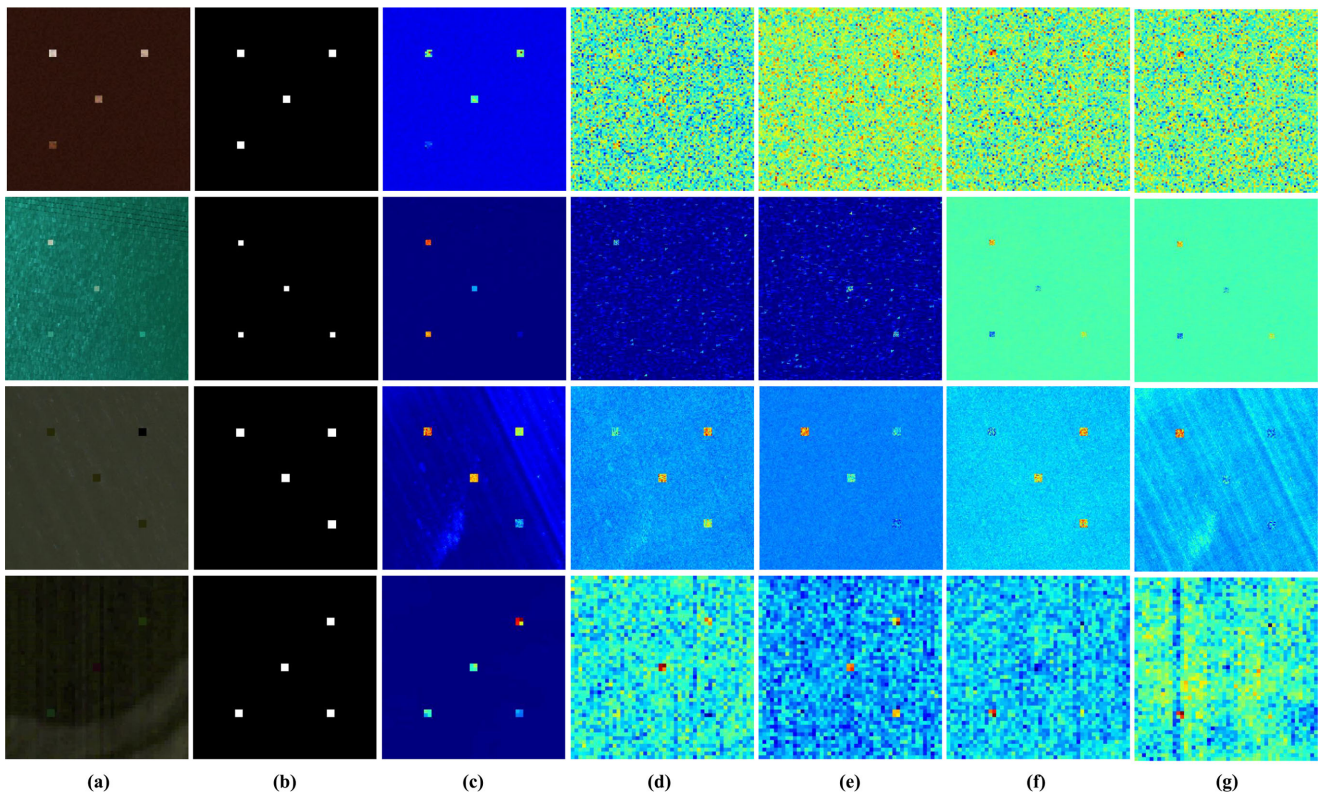


Fig. 10. Underwater target detection results of compared methods for the real datasets. (a) Color composites of synthetic, sea, river, and lake scenarios. (b) Reference map. (c) UTD-Net. (d) CEM. (e) MF. (f) UTDF. (g) GBF.

some corresponding experiments and results which are illustrated in Fig. 9, where target-water mixed pixels containing with different depth information are located in the synthetic underwater scene. The testing target-water mixed pixels consist of sheet metal material which is taken from USGS spectral library. From the Fig. 9, we can find that the target-water mixed pixels have been well detected under different depth conditions even though the target-water mixed spectrum is glued to the water background spectrum as shown in Fig. 9(a).

2) *Effectiveness Evaluation of Joint Anomaly Detector*: To verify whether the joint anomaly detection can better separate out the target-water mixed pixels from background, three classical anomaly target detectors are carried out for a comprehensive comparison. Owing to the special detect mechanism of the joint anomaly detector, underwater targets only locate at the identical depth in one testing dataset. For simplicity, we adopt the AUC values of anomaly detection results as the performance metric and the concrete results are exhibited in Table III. Apparently, the joint anomaly detector achieves a better anomaly detection accuracy indicating that this detector is capable of separating the target-water mixed pixels from the background robustly in different underwater scenarios.

3) *Effectiveness Evaluation of BMB Constraint*: To figure out whether the BMB constraint makes contribution to the target-water unmixing result, an autoencoder model trained without it has been performed on all the datasets. For the sake of representing the spectral distance, the Euclidean distance is utilized as a metric to measure target spectral distance $D\langle r_t(\lambda), \hat{r}_t(\lambda) \rangle$

and water spectra distance $D\langle r_\infty(\lambda), \hat{r}_\infty(\lambda) \rangle$. Table IV lists the detailed results while the characters (Y and N) in parentheses can be interpreted as whether the autoencoder has been trained with BMB constraint. According to Table IV, the spectra generated by the autoencoder with BMB constraint are more closer to the reference spectra. In other words, these spectra possess the desired physical meanings which guarantees the unmixing results make contribution to final underwater detection results. Furthermore, it is noteworthy that the results in the second and last columns of Table IV are the average values of many measurements which indicates that the unmixing results will not be stable if the BMB constraint is not applied to the autoencoder. Taken together, BMB constraint can help the autoencoder produce the stable unmixing results with specific physical meaning by acting as a training guidance.

D. Underwater Detection Performance

In this subsection, the underwater detection performance of the UTD-Net has been evaluated, generalized, and analyzed. Meanwhile, two prevalent underwater target detection algorithms GBF and UTDF are conducted for a comprehensive comparison. Besides, we also implement two widely used land-based target detection algorithms CEM and MF as the baselines.

For the datasets mentioned in previous subsection, the reference maps (such as scene pictures, ground truths, and underwater water detection maps) are exhibited in Fig. 10. Based on the visual inspection, our proposed method UTD-Net achieves the

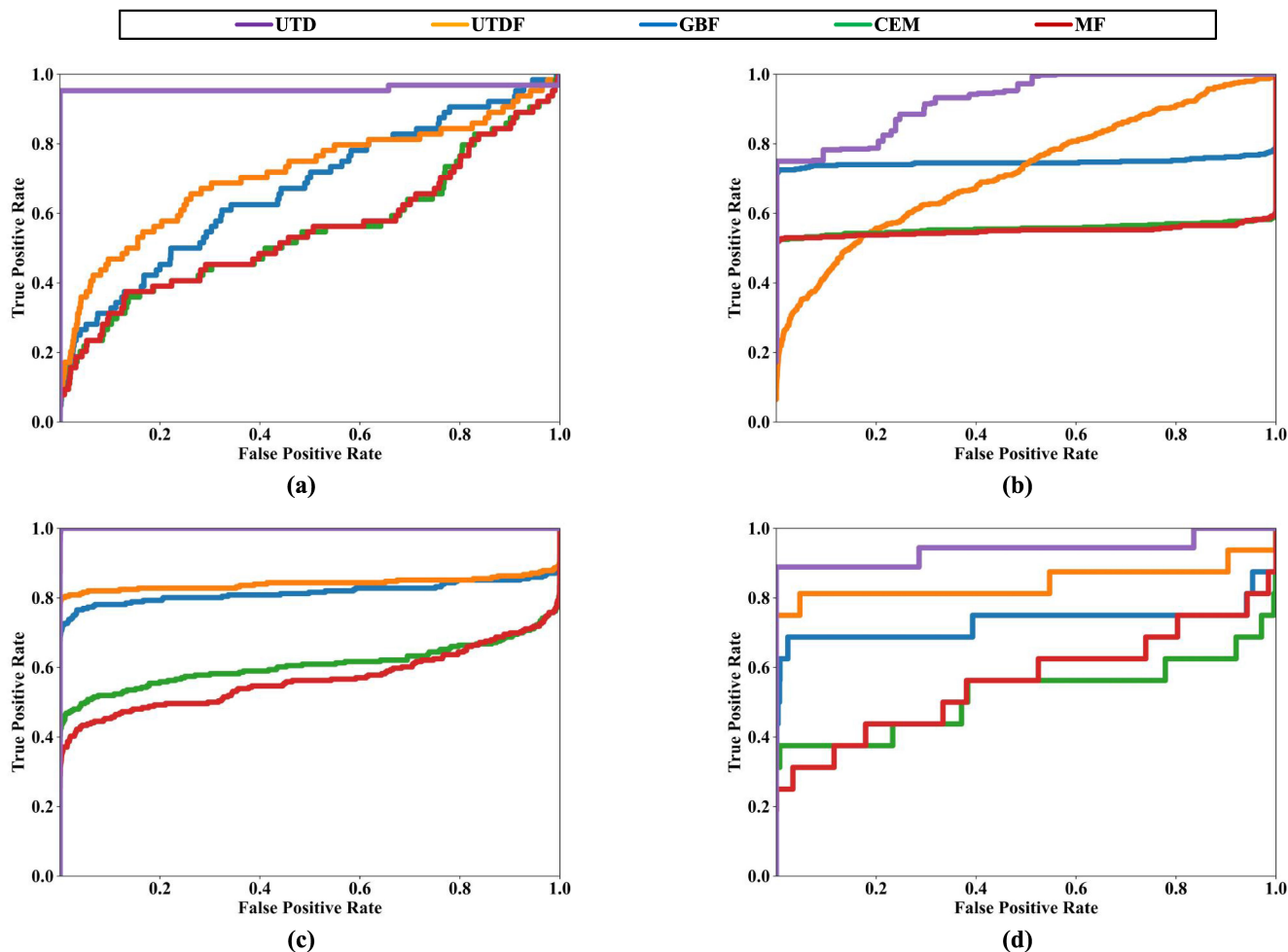


Fig. 11. ROC curves for the compared underwater target detection method. (a) Simulated dataset, (b) Sea dataset, (c) River dataset, and (d) Lake dataset.

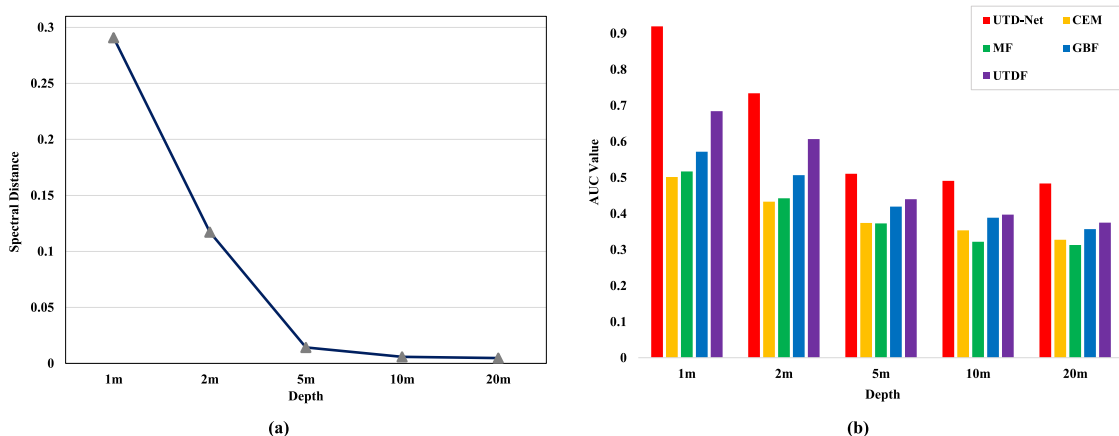


Fig. 12. Corresponding results for investigating the effect of depth information on synthetic water dataset. (a) Spectral distances between desired targets and background pixels with different depth information. (b) Detection performance of the compared methods with different depth information.

slightest visual difference with ground truths. This phenomenon accounts for the outperformance of UTD-Net in desired targets detection problem under various water environment conditions. From Fig. 10(d) to (g), we can also find that the water background has a great impact on the detection results for the compared

methods, which may lead to a higher false alarm rate in practical application. Obviously, only if the target spectra are diacritical can the compared methods finish the underwater targets detection task. On the contrary, the UTD-Net trends to suppress the background pixels while highlighting the desired underwater

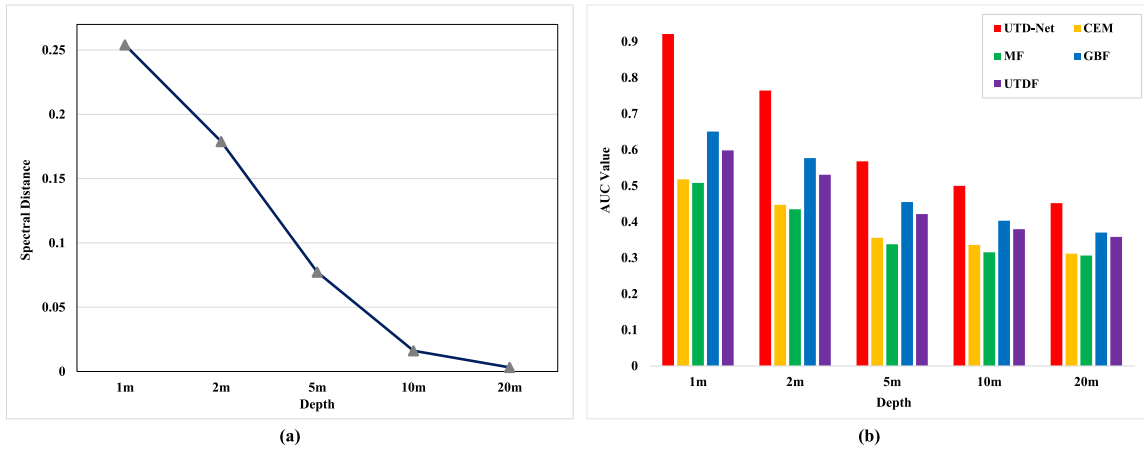


Fig. 13. Corresponding results for investigating the effect of depth information on sea water dataset. (a) Spectral distances between desired targets and background pixels with different depth information. (b) Detection performance of the compared methods with different depth information.

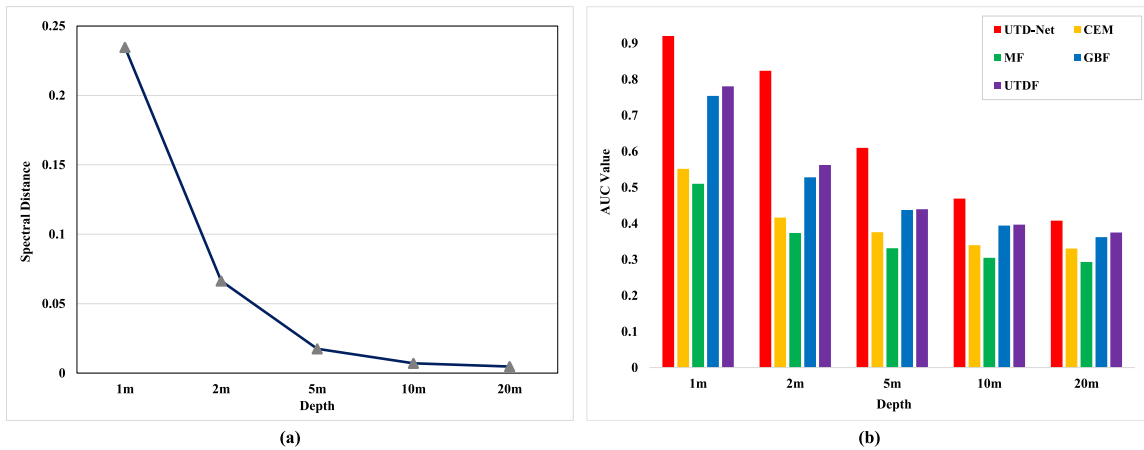


Fig. 14. Corresponding results for investigating the effect of depth information on river water dataset. (a) Spectral distances between desired targets and background pixels with different depth information. (b) Detection performance of the compared methods with different depth information.

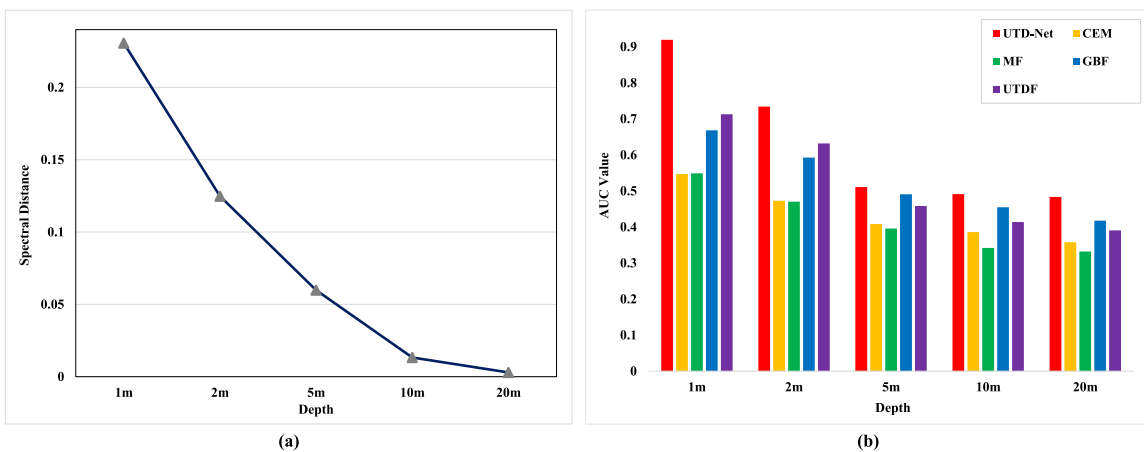


Fig. 15. Corresponding results for investigating the effect of depth information on lake water dataset. (a) Spectral distances between desired targets and background pixels with different depth information. (b) Detection performance of the compared methods with different depth information.

TABLE VI
AVERAGE COMPUTING TIME (IN SECONDS) OF THE COMPARED METHODS

Data set	Execution time (in seconds)				
	CEM	MF	GBF	UTDF	UTD-Net
Synthetic water	0.07272	0.02294	10.7841	15.8451	6.0798
Sea water	0.1276	0.3112	15.4782	24.1662	8.7447
River water	0.05086	0.1197	12.4119	19.4529	6.1176
Lake water	0.008924	0.004986	7.3684	11.4878	4.2826
Average	0.06502	0.1147	11.5017	17.7380	6.3062

The bold entries represent the best performance in each row.

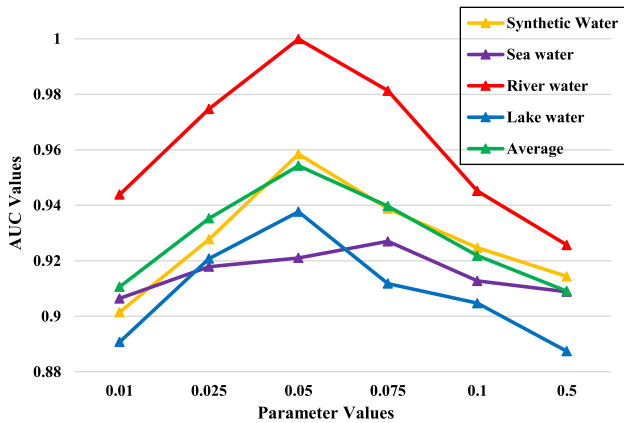


Fig. 16. Experimental results for the analysis of vital hyperparameter.

TABLE VII
STATISTICAL INFORMATION OF RESULTS FOR STABILITY ANALYSIS

Data set	Area under ROC values of UTD-Net			
	Before	Maximal	Minimal	Mean
Synthetic water	0.95848	0.9531	0.7738	0.8594
Sea water	0.92701	0.9249	0.7819	0.8341
River water	0.99995	0.9993	0.7023	0.8748
Lake water	0.93771	0.9362	0.7334	0.8341
Average	0.9558	0.9534	0.7479	0.8506

The label “Before” refers to the performance before imposing AWGN on datasets and the remainder items all represent statistical information after the datasets have been affected by AWGN.

targets pixels. Besides, owing to the pale depth estimation strategy, the compared methods cannot find out the targets in different depths simultaneously. As for UTD-Net, incorporating the spectral information of underwater targets shown in the Figs. 5 to 8 and the detection results, it is effortless to make the conclusion that our proposed method is competent to attain an approving detection performance even if the spectrum of underwater target is not so distinguishable from the surrounding water spectra. In other words, UTD-Net possesses the ability of seeking out the targets locating in a deeper position which will contribute a lot to deep-water exploration. Besides, UTD-Net can provide a stable and robust result regardless of the variability of target depth information.

In order to acquire a detailed qualitative analysis of the detection results, we plot the log-scale ROC curves for false alarm

rate P_f and target detection probability P_d . As demonstrated in Fig. 11, the proposed method remains over CEM, MF, GBF, and UTDF for all the datasets. This verifies that the UTD-Net can achieve a preferable underwater detection performance and maintain a lower false alarm rate simultaneously. For quantitative comparison, the AUC values of above ROC curves are also calculated and listed in Table V. Notably, UTD-Net has still achieved the optimal results for all the datasets. Moreover, the AUC values of our proposed method exceed 0.9 in all the underwater scenarios. In particular, the AUC value in the river water dataset is approximate to the optimal result 1. Furthermore, the average AUC score of UTD is 0.9558 while the suboptimum performance of compared methods is merely 0.7792, which shows the super performance of our proposed method in a numerical aspect.

After that, we also measure the effect of depth information on the final detection results. For simplicity, the desired targets contained in one dataset would be positioned at the identical position and the AUC values are exploited as the metric. To roundly exhibit the measurement results, we tune the depth information of desired targets in each dataset from 1 to 20 m with an alterable step while Figs. 12 to 15 are devoted to demonstrating the corresponding detection performances. We first calculate the spectral distance between desired target and water column as depth information alters for all the datasets, where the line charts plotted in Figs. 12(a) to 15(a) are utilized to depict this specific relationship. Obviously, this specific distance decreases with the growth of depth information, which makes it more difficult to detect desired targets since they are indistinguishable with the background water column. As a result, the detection performances of all the compared methods suffers if we would like to detect the deeper desired targets. It is easy to find that UTD-Net can still achieve the best detection performances under all the appointed depth information although the performance gaps with other testing methods are shrinking. However, even if the depth information are set as 20 m, UTD-Net can still surpass the baselines (CEM and MF) with depth information 2 m in all the datasets. This confirms that our proposed method is capable of contributing to underwater detection tasks in spite of desired targets locating in the deep positions.

Finally, the execution time of different methods is also exploited as another criterion to offer quantitative analysis. All the tests are carried out on the same experimental condition mentioned in the subsection *Experiment Setting*. It is evident that it takes a quantity of time to train the model for a learning-based method. Consequently, we only measure the running time of detection stage when the model training has been completed. As listed in Table VI, the computation consumption of our proposed method is not the optimal, but it is less than GBF and UTDF. CEM and MF have the similar computational complexity which surpass other compared methods. However, from the detection performance we can know that these two methods cannot obtain the satisfying detection results. The reason accounting for their less computational complexity is that land-based detection methods do not take the impact of surrounding water environment into consideration. In terms of UTD-Net, it can achieve the optimal underwater detection result with a tolerable time consumption.

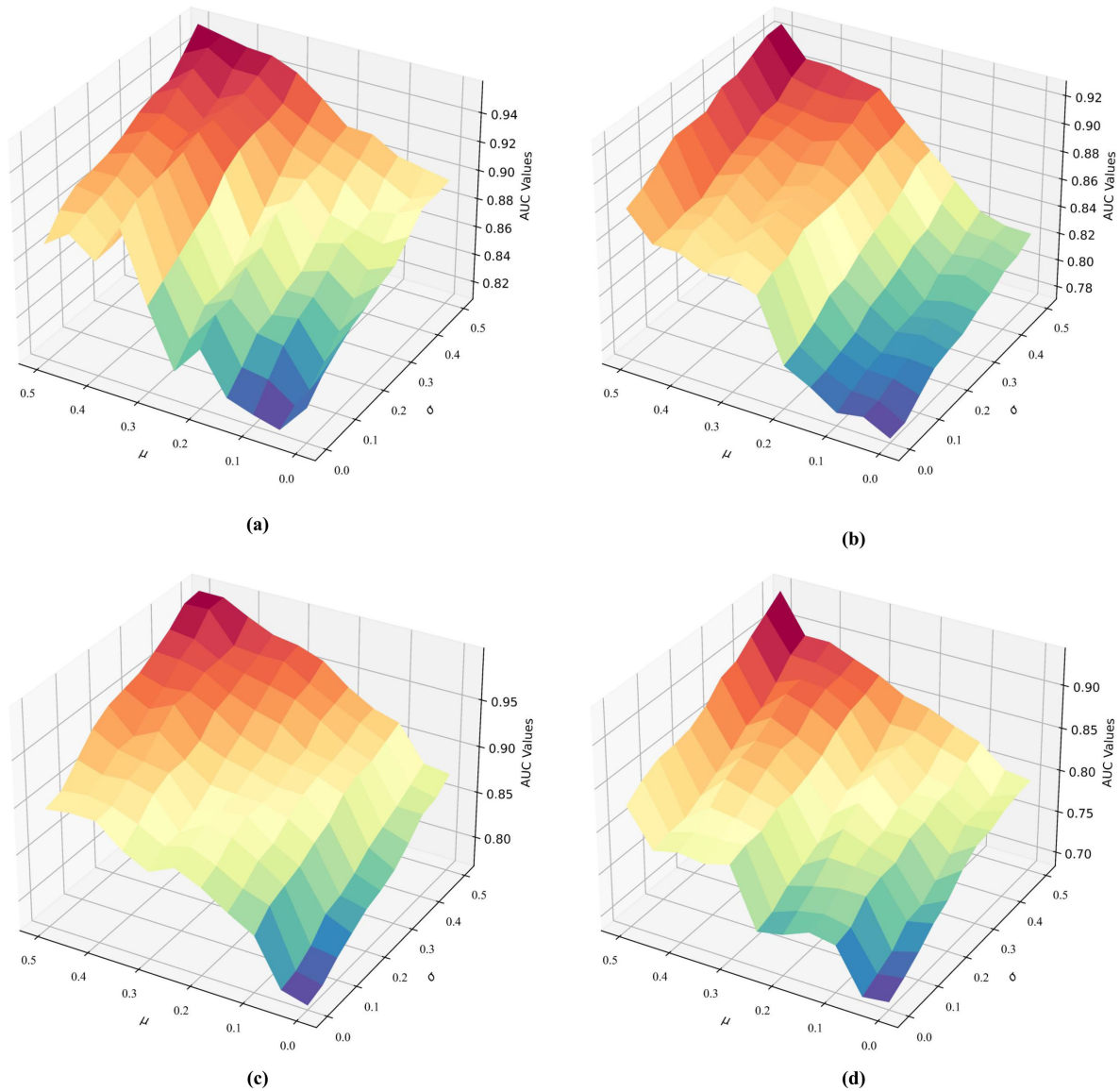


Fig. 17. AUC curves for UTD-Net under different noisy conditions. (a) Synthetic dataset, (b) Sea dataset, (c) River dataset, and (d) Lake dataset.

Summarizing every aspect of the experiment performances, we can finally come to the conclusion that UTD-Net possesses the capacity to achieve a promising underwater detection performance in most underwater scenarios.

E. Method Analysis

In this section, we will perform some special experiments to comprehensively demonstrate the proposed method in terms of hyperparameter analysis, stability analysis, and target size analysis.

1) *Analysis of Vital Hyperparameter*: The function mentioned in (20) refers to the objective function of UTD-Net and the parameter λ is exploited to establish a tradeoff between reconstruction error and BMB constraint. Consequently, it is necessary to verify whether the detection results are sensitive to this vital parameter λ . To accomplish this task, we conduct UTD-Net with different λ values on all the datasets and a line chart is established

to record the corresponding results in Fig. 16. From this chart we can find that the detection performances will meet the maximum values while enlarging the value of hyperparameter λ . Besides, 0.05 is the maximum points for synthetic water, river water, and lake water datasets while 0.075 is the maximum point for sea water dataset. Therefore, hyperparameter λ could have impact on the final detection results and different datasets would possess different maximum points of this hyperparameter. Moreover, according to the average result in Fig. 16, UTD-Net achieves almost the same detection performances when the value of λ are 0.01 and 0.5. Meanwhile, the performance gap between the best λ value and the worst λ value is less than 0.05. These phenomena confirm that UTD-Net is not sensitive to the hyperparameter λ but a relatively small value is recommended in practical application for a more favorable convergence speediness.

2) *Analysis of Method Stability*: There is no doubt that the effects of noise always exist in practice and an excellent underwater detection method should achieve the remarkable results

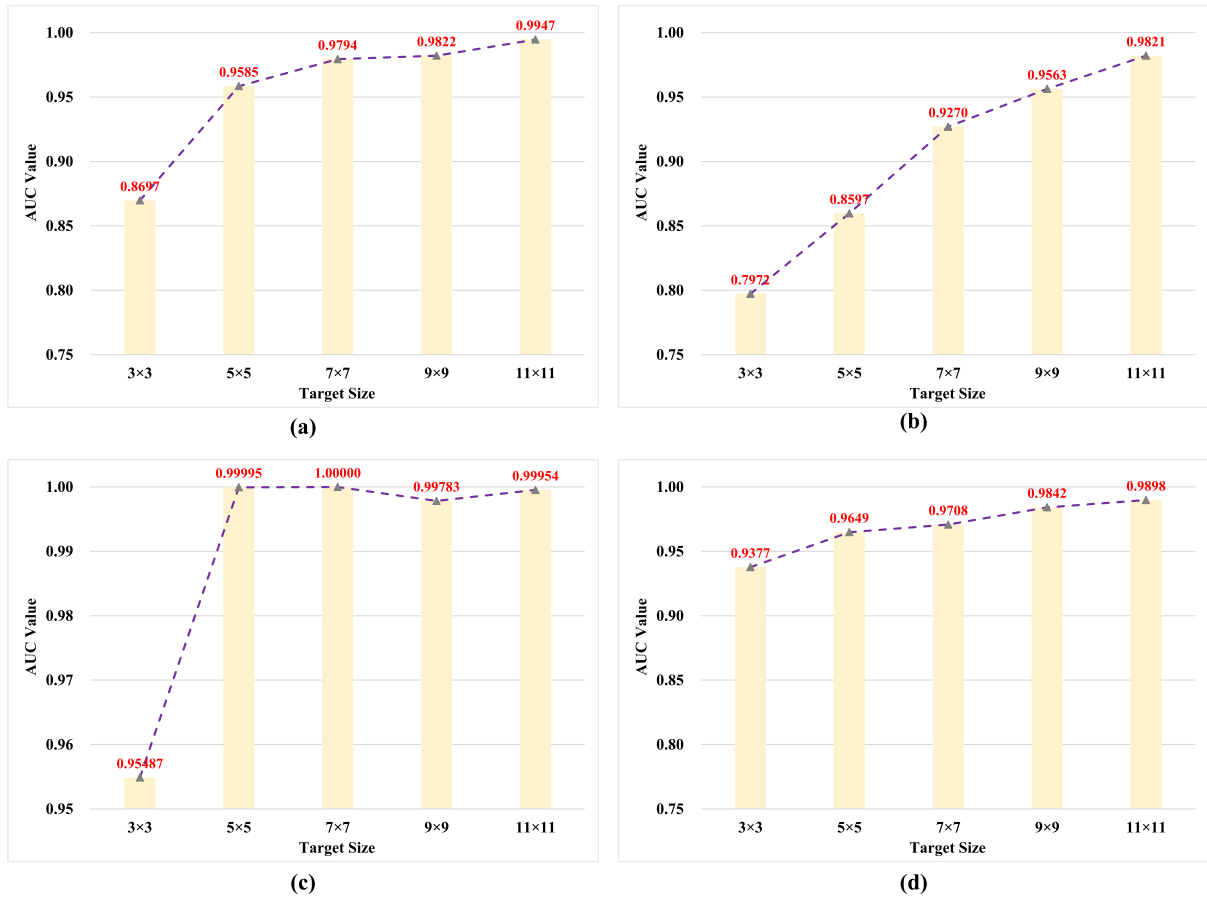


Fig. 18. Detection performances of UTD-Net with different target sizes. (a) Synthetic dataset, (b) Sea dataset, (c) River dataset, and (d) Lake dataset.

regardless of the interference of noise. In order to analyze the influence of noise on final detection results, an additive white Gaussian noise (AWGN) is imposed on the all the datasets which can be formulated as follows:

$$n(\lambda) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\lambda - \mu)^2}{2\sigma^2}\right) \quad (25)$$

where μ and σ refers to mean vector and covariance metric. Note that, the mean value μ and covariance vaule σ in different bands would be endowed with the same values and then these two parameters come to be the only variables for AWGN. Therefore, we tune the values of μ and σ to generate different AWGNs whlie the 3D figures are employed to record the corresponding detection results, which are demonstrated in Fig. 17. On the basis of above 3D plots, we can get to know that the UTD-Net is more sensitive to mean value and the AWGN cannot bring vast adverse influence to the final detection performances with the assistance of visual inspection. The relevant statistical information has been collected in Table VII for the quantitative analysis, where the label “before” refers to the performance before imposing AWGN on datasets and the remainder items all represent statistical information after the datasets have been affected by AWGN. Integrating the Tables V and VII, we can find that the worst performances of UTD-Net with AWGN can still surpass the other compared methods in synthetic water and sea water datasets. Meanwhile, the gaps between worst

detection performances and the performances before inflicted by AWGN are not so distinct that they can be ignored in practical application. To make a summary, UTD-Net is capable of acting as a stable underwater detection method when suffering from the interference of AWGN.

3) *Analysis of Target Size and Shape*: Target size and target shape are two important parameters for the experimental datasets. Since UTD-Net turns out to be an unmixing-based underwater detection method, the proportion of target pixels may affect the final detection performance. Moreover, target size is the crucial parameter to control the proportion of target pixels when the spatial solution of experimental chip has been determinate. To research the impact of target size on the final detection accuracy, we tune the sizes of desired targets mentioned in Section IV-A from 3×3 to 11×11 for all the datasets and perform UTD-Net to accomplish the underwater target detection missions. The associate detection results are displayed in the Fig. 18. The polylines in above figures reflect the change trends of detection accuracies with the growth of target sizes. Obviously, a larger target size would contribute to improving the detection accuracy since it enlarges the energy of the desired targets and makes them more distinguishable from background. This effect will make it easier for UTD-Net to separate the target pixels from background with endmembers extraction scheme. However, desired targets with large target size are not common in practical application while dim and weak target detection

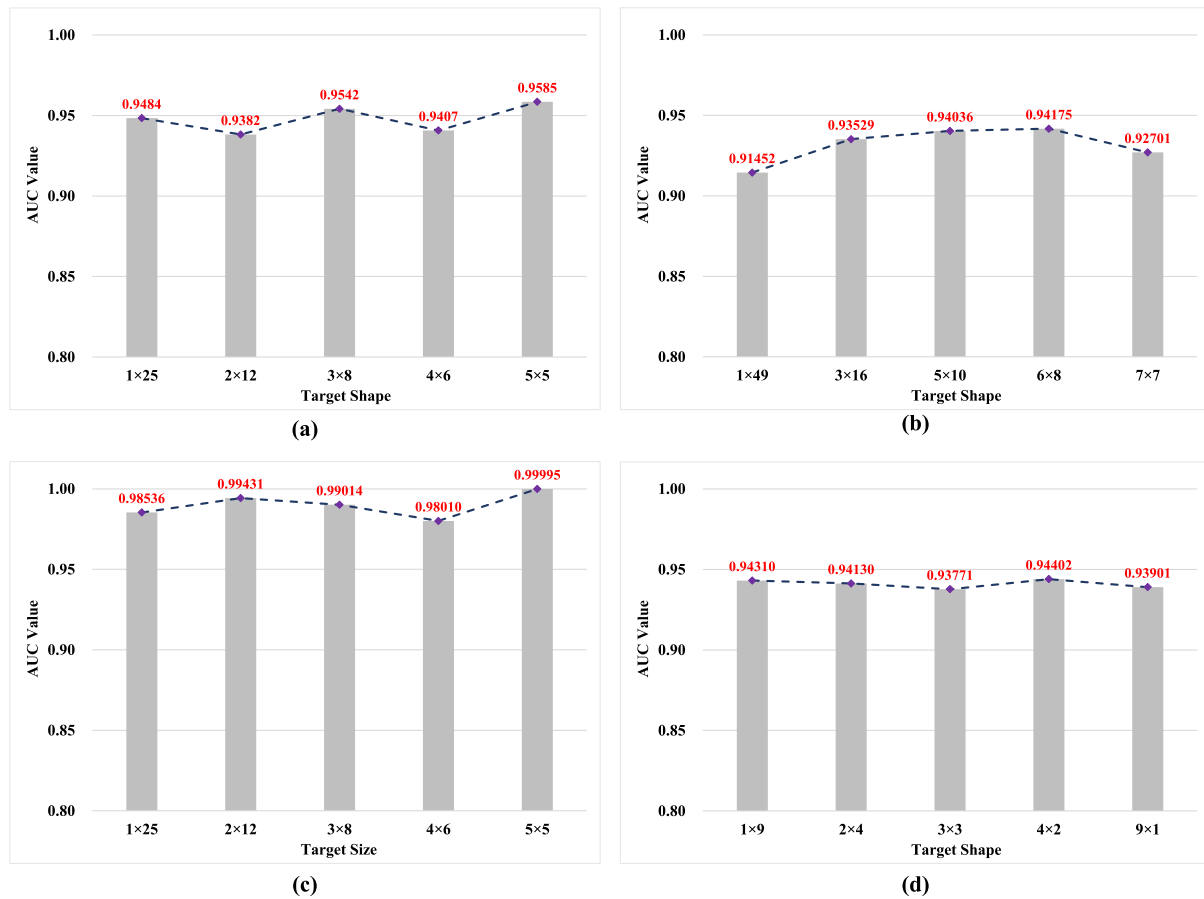


Fig. 19. Detection performances of UTD-Net with different target shapes on (a) Synthetic dataset, (b) Sea dataset, (c) River dataset, and (d) Lake dataset.

issue remains a big challenge in the hyperspectral detection filed.

As for the influence of target shape, we decide to change the shapes of desired targets under the condition that the total amounts of target pixel remains almost the same. Fig. 19 is exploited to demonstrate the corresponding results. According to the qualitative and quantitative judgements, the target shape would not bring dramatic change to the detection performance and the variety of detection results among different target shape might derive from experiment errors. The reason accounting for this phenomenon is that UTD-Net detects the desired target based on unmixing methodology instead of employing the contextual information of desired targets. Consequently, the target shapes might not have impact on the detection accuracy.

V. CONCLUSION

In this article, an unmixing based detection network named UTD-Net has been proposed to separate the desired underwater targets from water background. More precisely, UTD-Net consists of two meaningful parts: Joint anomaly detector and BMB autoencoder. Due to the interference of background pixels on final detection result, we develop a joint anomaly detector to separate out the target-water mixed pixels as anomalies while suppressing the effect of background. Considering the high performance of the HU method in anomaly detection problems, the

joint anomaly detector is designed based on three classical HU methods. A fusion strategy called CBAA has also been designed to attain a more robust and effective anomaly detection result. Then, for the sake of unmixing target-spectral mixed pixels, we present a novel deep learning-based autoencoder with specific constraint. In the encoder part of above autoencoder, 1D CNN structures are exploited to construct a data transformer which can compress the original spectral data into the abundance vectors. In this way, we manage to predict the abundance values of unmixing result while addressing the spectral variability problem raised by environment factors. In addition, to tackle the issue of physically meaningless endmembers in blind unmixing, the BMB constraint is imposed on the objective function as a guidance to assist autoencoder in generating desired unmixing results. Finally, based on the physical essence of abundance maps, the underwater target detection result is derived by fusing outcomes of BMB autoencoder with weight coefficients determined by the abundance values. Experiments carried out on several synthetic sets show that UTD-Net achieves excellent performances for underwater target detection problems and outperforms the prevalent methods in both objective and subjective evaluation.

ACKNOWLEDGMENT

The authors would like to thank the R&D team of Gaofen-5 Visible short-wave infrared hyperspectral camera of Shanghai

Institute of Technical Physics, Chinese Academy of Sciences for providing the HSIs obtained by Gaofen-5 ASHI. AHSI is one of the most advanced payloads, it has a wider swath, higher spectral resolution, and higher signal-to-noise ratio, showing excellent performance in underwater target detection and providing strong support for our research work.

REFERENCES

- [1] S. Mei, G. Zhang, J. Li, Y. Zhang, and Q. Du, "Improving spectral-based endmember finding by exploring spatial context for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3336–3349, Jun. 2020.
- [2] X. Xu, J. Li, S. Li, and A. Plaza, "Curvelet transform domain-based sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4908–4924, Aug. 2020.
- [3] L. He, J. Zhu, J. Li, D. Meng, J. Chanussot, and A. Plaza, "Spectral-fidelity convolutional neural networks for hyperspectral pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5898–5914, Sep. 2020.
- [4] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, Jan. 2018.
- [5] D. Zhu, B. Du, and L. Zhang, "Target dictionary construction-based sparse representation hyperspectral target detection methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1254–1264, Apr. 2019.
- [6] W. Xie, X. Zhang, Y. Li, K. Wang, and Q. Du, "Background learning based on target suppression constraint for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5887–5897, Sep. 2020.
- [7] Y. Zhang, K. Wu, B. Du, and X. Hu, "Multitask learning-based reliability analysis for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2135–2147, Feb. 2019.
- [8] W. Xie, J. Yang, J. Lei, Y. Li, Q. Du, and G. He, "SRUN: Spectral regularized unsupervised networks for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1463–1474, Oct. 2020.
- [9] R. Amin, R. Gould, W. Hou, R. Arnone, and Z. Lee, "Optical algorithm for cloud shadow detection over water," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 732–741, Feb. 2013.
- [10] Z. Haigang, L. Zhongping, S. Ping, C. Chuqun, and K. L. Carder, "Retrieval of water optical properties for optically deep waters using genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1123–1128, May 2003.
- [11] J. G. P. W. Clevers and L. Kooistra, "Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 574–583, Apr. 2012.
- [12] C. D. Mobley, "Comparison of numerical models for computing underwater light fields," *Appl. Opt.*, vol. 32, no. 36, pp. 7484–7504, 1993.
- [13] P. J. Werdell *et al.*, "An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing," *Prog. Oceanogr.*, vol. 160, pp. 186–212, 2018.
- [14] R. J. W. Brewin, E. Devred, S. Sathyendranath, S. J. Lavender, and N. J. Hardman-Mountford, "Model of phytoplankton absorption based on three size classes," *Appl. Opt.*, vol. 50, no. 22, pp. 4535–4549, 2011.
- [15] Q. Dong, S. Shang, and Z. Lee, "An algorithm to retrieve absorption coefficient of chromophoric dissolved organic matter from ocean color," *Remote Sens. Environ.*, vol. 128, pp. 259–267, 2013.
- [16] S. Jay, M. Guillaume, and J. Blanc-Talon, "Underwater target detection with hyperspectral data: Solutions for both known and unknown water quality," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1213–1221, Aug. 2012.
- [17] D. B. Gillis, "An underwater target detection framework for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1798–1810, Apr. 2020.
- [18] D. B. Gillis, J. H. Bowles, M. J. Montes, and W. D. Miller, "Deriving bathymetry and water properties from hyperspectral imagery by spectral matching using a full radiative transfer model," *Remote Sens. Lett.*, vol. 11, no. 10, pp. 903–912, 2020.
- [19] A. Albert and C. D. Mobley, "An analytical model for subsurface irradiance and remote sensing reflectance in deep and shallow case-2 waters," *Opt. Exp.*, vol. 11, no. 22, pp. 2873–2890, 2003.
- [20] Z. Lee, K. L. Carder, C. D. Mobley, R. G. Steward, and J. S. Patch, "Hyperspectral remote sensing for shallow waters. I. A semi-analytical model," *Appl. Opt.*, vol. 37, no. 27, pp. 6329–6338, 1998.
- [21] Z. Lee, K. L. Carder, C. D. Mobley, R. G. Steward, and J. S. Patch, "Hyperspectral remote sensing for shallow waters: Deriving bottom depths and water properties by optimization," *Appl. Opt.*, vol. 38, no. 18, pp. 3831–3843, 1999.
- [22] R. Clark *et al.*, "USGS digital spectral library splib06a," Data Series 231, U.S. Geological Surv., USA, 2007. [Online]. Available: <http://speclab.cr.usgs.gov/spectral-lib.html>.
- [23] R. A. Garcia, L. I. W. McKinna, J. D. Hedley, and P. R. C. S. Fearn, "Improving the optimization solution for a semi-analytical shallow water inversion model in the presence of spectrally correlated noise," *Limnol. Oceanogr. Methods*, vol. 12, no. 10, pp. 651–669, 2014.
- [24] J. Hedley, C. Roelfsema, and S. R. Phinn, "Efficient radiative transfer model inversion for remote sensing applications," *Remote Sens. Environ.*, vol. 113, no. 11, pp. 2527–2532, 2009.
- [25] Y. Zhang *et al.*, "An improved low rank and sparse matrix decomposition-based anomaly target detection algorithm for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2663–2672, May 2020.
- [26] C. Zhao, C. Li, S. Feng, N. Su, and W. Li, "A spectral-spatial anomaly target detection method based on fractional fourier transform and saliency weighted collaborative representation for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5982–5997, Oct. 2020.
- [27] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, May 2012.
- [28] B. Yang, Z. Chen, and B. Wang, "Nonlinear endmember identification for hyperspectral imagery via hyperpath-based simplex growing and fuzzy assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 351–366, Jan. 2020.
- [29] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravorty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, Jan. 2019.
- [30] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, May 2014.
- [31] N. B. Karayiannis, "An axiomatic approach to soft learning vector quantization and clustering," *IEEE Trans. Neural Netw.* vol. 10, no. 5, pp. 1153–1165, Sep. 1999.
- [32] C. Kwan, B. Ayhan, G. Chen, W. Jing, J. Baohong, and I. C. Chein, "A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 409–419, Jan. 2006.
- [33] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.
- [34] H. Su, Z. Wu, Q. Du, and P. Du, "Hyperspectral anomaly detection using collaborative representation with outlier removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5029–5038, Dec. 2018.
- [35] R. Richter and D. Schlöpfer, "Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction," *Int. J. Remote Sens.*, vol. 23, no. 13, pp. 2631–2649, 2002.
- [36] J. W. Boardman, F. A. Kruse and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data", in *Proc. JPL Airborne Earth Sci. Workshop*, 1995, pp. 23–26.
- [37] M. E. Winter, "N-finder: An algorithm for fast autonomous spectral endmember determination in hyperspectral data". *Proc. SPIE*, vol. 3753, pp. 266–277, 1999.
- [38] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
- [39] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, Univ. Maryland, College Park, MD, USA, 1993.
- [40] D. Manolakis and G. Shaw, "Directionally constrained or constrained energy minimization adaptive matched filter: Theory and practice," *Proc. SPIE*, vol. 4480, Aug. 2001.

- [41] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [42] I. C. Chein, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Feb. 2005.
- [43] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerosp. Conf.*, Mar. 2004, pp. 1818–1824.



Jiahao Qi received the bachelor's degree in electronic and information engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2019. He is currently working toward the master's degree in information and communication engineering from the National Key Laboratory of Science and Technology on ATR, National University of Defense Technology, Changsha, China.

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



Zhiqiang Gong received the bachelor's degree in applied mathematics from Shanghai Jiao Tong University, Shanghai, China, in 2013, the master's degree in applied mathematics from the National University of Defense Technology (NUDT), Changsha, China, in 2015, and the Ph.D. degree in information and communication engineering from the National Key Laboratory of Science and Technology on ATR, NUDT, in 2019.

He is currently an Assistant Professor with the National Innovation Institute of Defense Technology, Chinese Academy of Military Science, Beijing, China. He has authored more than ten peer-reviewed articles in international journals, such as the *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, the *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, the *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, and the *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*. His research interests include computer vision, machine learning, and image analysis.



Wei Xue received the Ph.D. degree in computer science and technology from the Nanjing University of Science and Technology, Nanjing, China, in 2017.

From July 2014 to April 2017, he was a Visiting Student with the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently a Lecturer with the School of Computer Science and Technology, Anhui University of Technology, Maanshan, China, and also is a Postdoctoral Researcher with the National Key Laboratory of Science and

Technology on Automatic Target Recognition, National University of Defense Technology, Changsha, China. His research interests include machine learning, computer version, and nonlinear optimization.



Xingyue Liu received the bachelor's degree in communication engineering from Southwest University, Chongqing, China, in 2018, the master's degree in applied information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2020. She is currently working toward the doctor's degree in information and communication engineering from National University of Defense Technology, Changsha, China.

Her research interests include multisensor data fusion, image processing, computer vision, and machine learning.



Aihuan Yao received the bachelor's degree in electronic engineering from Xidian University, Xi'an, China, in 2019. She is currently working toward the master's degree in information and communication engineering from the National Key Laboratory of Science and Technology on ATR, National University of Defense Technology, Changsha, China.

Her interests include remote sensing image processing and machine learning.



Ping Zhong (Senior Member, IEEE) 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 March 2015 to February 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 National Key Laboratory of Science and Technology on ATR, NUDT. He has authored more than 40 peer-reviewed articles 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*, the *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING*, and the *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*. His research interests include computer vision, machine learning, and pattern recognition.

Dr. Zhong was the recipient of the National Excellent Doctoral Dissertation Award of China, in 2011, and the New Century Excellent Talents in the University of China, in 2013. He is also a Referee for 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 the *IEEE Geoscience and Remote Sensing Letters*.