remote sensing

MDPI

Article

Target Detection Adapting to Spectral Variability in Multi-Temporal Hyperspectral Images Using Implicit Contrastive Learning

Xiaodian Zhang [1], Kun Gao [1,*], Junwei Wang [1], Pengyu Wang [1], Zibo Hu [1], Zhijia Yang [1], Xiaobin Zhao [2] and Wei Li [2]

[1] Key Laboratory of Photoelectronic Imaging Technology and System, Beijing Institute of Technology, Beijing 100081, China; xdz@bit.edu.cn (X.Z.); wjw@bit.edu.cn (J.W.); 3120210542@bit.edu.cn (P.W.); zibohu@bit.edu.cn (Z.H.); 3120225323@bit.edu.cn (Z.Y.)

[2] School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China; xiaobinzhao@bit.edu.cn (X.Z.); liw@bit.edu.cn (W.L.)

* Correspondence: gaokun@bit.edu.cn

Abstract: Hyperspectral target detection (HTD) is a crucial aspect of remote sensing applications, aiming to identify targets in hyperspectral images (HSIs) based on their known prior spectral signatures. However, the spectral variability resulting from various imaging conditions in multi-temporal hyperspectral images poses a challenge to both classical and deep learning (DL) methods. To overcome the limitations imposed by spectral variability, an implicit contrastive learning-based target detector (ICLTD) is proposed to exploit in-scene spectra in an unsupervised way. First, only prior spectra are utilized for explicit supervision, while an implicit contrastive learning module (ICLM) is designed to normalize the feature distributions of prior and in-scene spectra. This paper theoretically demonstrates that the ICLM can transfer the gradients from prior spectral features to those of in-scene spectra based on their feature similarities and differences. Because of transferred gradient signals, the ICLTD is regularized to extract similar representations for the prior and in-scene target spectra, while augmenting feature differences between the target and background spectra. Additionally, a local spectral similarity constraint (LSSC) is proposed to enhance the capability of scene adaptation by leveraging the spectral similarities among in-scene targets. To validate the performance of the ICLTD under spectral variability, multi-temporal HSIs captured under various imaging conditions are collected to generate prior spectra and in-scene spectra. Comparative evaluations against several DL detectors and classical methods reveal the superior performance of the ICLTD in achieving a balance between target detectability and background suppressibility under spectral variability.

Keywords: hyperspectral target detection; remote sensing; spectral variability; multi-temporal hyperspectral images

check for updates

Citation: Zhang, X.; Gao, K.; Wang, J.; Wang, P.; Hu, Z.; Yang, Z.; Zhao, X.; Li, W. Target Detection Adapting to Spectral Variability in Multi-Temporal Hyperspectral Images Using Implicit Contrastive Learning. *Remote Sens.* **2024**, *16*, 718. https://doi.org/10.3390/rs16040718

Academic Editor: Salah Bourennane

Received: 24 December 2023
Revised: 13 February 2024
Accepted: 16 February 2024
Published: 18 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the advancement of hyperspectral imaging techniques, the spectral and spatial characteristics of hyperspectral images (HSIs) have found applications in remote sensing observations, including military defense [1], mineral exploration [2], and agricultural monitoring [3]. Hyperspectral target detection (HTD), which stands out as a critical technique for interpreting remote sensing HSIs, aims to identify the target of interest within the test HSIs given their known waveform signatures. This paper focuses on HTD in the radiance domain where the prior spectra and test HSIs are collected from multi-temporal images. For clarity of description, "in-scene HSIs" refers to the test images in this paper. An example shown is in Figure 1, where the diverse imaging conditions of the multi-temporal HSIs lead to variations between the prior and in-scene target spectra, known as spectral variability [4]. This variability presents a significant obstacle to achieving robust

detection performance. Although the in-scene radiative parameters could be potentially estimated to address radiance variations [5], such estimation necessitates the presence of typical materials in the in-scene HSIs and prior knowledge of their spectra. This work aims to mitigate the limitations posed by spectral variability and realize robust detection performance without prior knowledge of radiative parameters.

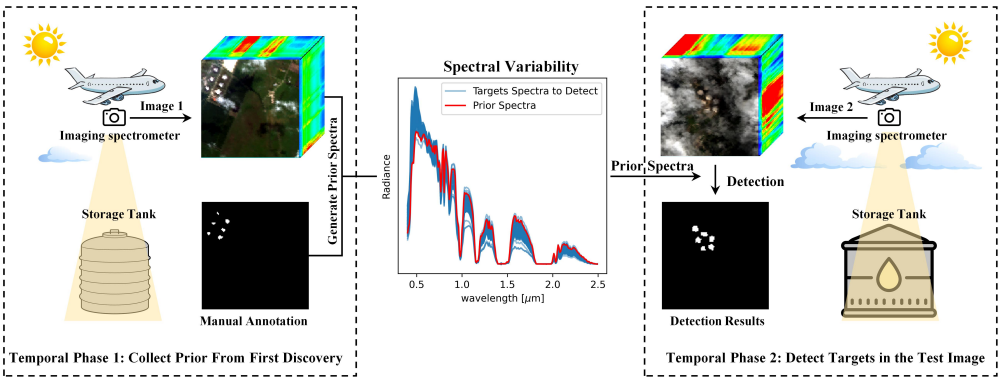

Figure 1. Pipeline of HTD where prior spectra and the test spectra are captured in different time phases. The prior spectra could either be obtained from a spectral library or known target spectra from other HSIs (the latter being the focus of this paper). The prior spectra differ from the target spectra due to various imaging conditions and intra-class variance, known as spectral variability, which poses challenges to HTD.

In the last few decades, numerous classical methods have been developed for HTD. Spectral angle mapper [6] and spectral information divergence [7] are two well-known distance-based metrics used to assess the similarities between prior and in-scene spectra. Signal detection-based approaches, such as the generalized likelihood ratio test [8] and the adaptive coherence/cosine estimator (ACE) [9,10], employ hypothesis tests to model HTD. Constrained energy minimization (CEM) [11] can be used to construct a detector that minimizes the output energy while preserving prior spectra. To overcome the limitations of the distribution assumption, Chang et al. [12] proposed more general theories based on the signal-to-noise ratio and spectral angle, including the spectral angle in the correlated space (R-SA). In addition, innovative approaches have been integrated into classical methods to address limitations arising from model complexity. These include kernel methods [13], hierarchical structures [14–16], and fractional Fourier transforms [17]. To utilize in-scene spectra better, spectral unmixing techniques [18,19] and sparsity assumptions [20–24] are introduced to construct hyperspectral target detectors, which require certain assumptions.

In the past decade, deep learning (DL) models have achieved great success in HSI interpretation based on their complex model representation capabilities, such as classification [25], mapping [26], and unmixing [27]. Meanwhile, DL-based HTD approaches have also received considerable attention. Supervised learning-based detectors were first proposed to optimize the network in a contrastive learning method [28]. Two-stream [29,30] and Siamese structures [31–33] are commonly used to realize contrastive learning. With the advancement of neural networks, there have been significant improvements in feature extraction techniques employed by supervised learning approaches. These techniques have evolved from fully connected networks [28] and convolutional networks [29] to more recent transformer models [31,34]. The optimization of numerous network parameters requires abundant labeled data, while only prior spectra are available with annotations. Therefore, supervised HTD methods commonly generate pseudo-data (e.g., pre-detect spectra from in-scene HSIs or simulate spectra) to supplement the training data. Classical detectors [30,31], end-member extraction [31], and clustering methods [29] are employed to pre-detect background samples from in-scene HSIs. Background samples that are easily misidentified as targets are valuable for training; however, they are relatively difficult to detect. Spectral mixture models [29,32,35,36] and generative adversarial networks [37,38] are

mainstream simulation approaches. However, simulating high-quality spectra is hard under spectral variability. For supervised learning-based detectors, the reliability of generated pseudo-data limits detection performance.

To overcome the few-shot samples problem, unsupervised and self-supervised learning-based detectors were proposed to optimize networks with pretext tasks [39], which are free of manual annotations. Typical pretext tasks for HTD include spectra reconstruction [40–42], denoising [43], and band selection [44]. Ref. [38] proposed a novel pretext task that constructs contrastive learning between odd and even bands of unlabeled spectra. These pretext tasks allow the detectors to learn the intrinsic properties of the HSIs and extract representations of the in-scene spectra. Because pretext tasks do not introduce supervision from manual annotations, supervised fine-tuning [45] or regularization [46–48] is embedded in the optimization of pretext tasks, which commonly also requires labeled data. Alternatively, additional classifiers or post-processing are necessary for the final results [41,49,50], which complicates the detection process.

To summarize, exploiting in-scene spectra is crucial to alleviate the challenges posed by spectral variability. Non-learning methods model in-scene spectra under certain assumptions, such as statistical assumptions and spectral mixing models. For DL methods, utilizing in-scene spectra to generate pseudo-data is commonly adopted for supervised learning or fine-tuning. However, ensuring the quality and reliability of the generated spectra-annotation pairs requires significant effort. Additionally, most current approaches focus on HTD using single-epoch HSIs, where prior spectra and in-scene spectra are captured during the same time frame. Dealing with spectral variations of targets in multi-temporal images is more challenging.

Unlike pseudo-data-based detectors, which distinguish background and target spectral samples under the guidance of numerous pseudo annotations, this paper proposes a pseudo-data-free detector that exploits in-scene spectra with implicit contrastive learning [51], which realizes contrastive learning without manual supervision. The proposed detector is explicitly optimized to classify prior spectra while no other annotations are needed. To prevent model collapse resulting from few-shot training samples, an implicit contrastive learning module (ICLM) is proposed to regularize the detector to minimize the loss function through learning differentiated representations of in-scene spectra instead of over-fitting the prior spectra. Specifically, the ICLM normalizes the latent features of prior and in-scene spectra and establishes gradient propagation paths between them. Later in this paper, we theoretically analyze how the ICLM allocates differentiated gradient signals from prior spectral features to the latent features of unlabeled in-scene spectra based on their inherent feature differences. Based on the various allocated gradients, feature differences between targets and backgrounds in the in-scene HSIs are augmented during optimization. In addition to the ICLM, a local spectral similarity constraint (LSSC), which utilizes the spectral similarities of in-scene targets within local neighborhoods, is proposed to improve in-scene adaptability. Specifically, the LSSC assumes that the in-scene target spectra of multi-pixel targets are similar and spatially connected. To better study the detection performance under spectral variability, three multi-temporal HSI pairs collected from Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) data were used to conduct experiments. Each pair consisted of two HSIs captured under different imagining conditions, one for prior spectra generation and the other for target rediscovery experiments. The same category of targets appeared in each multi-temporal HSI pair. Four classical methods and three DL detectors were used for performance comparison under spectral variability.

The main contributions of this paper are as follows:

(1) This paper proposes a DL-based detector that adapts to spectral variability in multi-temporal hyperspectral images using implicit contrastive learning with prior and in-scene spectra.
(2) The ICLM is designed to regularize the optimization process of classifying prior spectra. This regularization helps in learning distinct representations of in-scene spectra based on their inherent differences from the prior spectra.

(3) The LSSC is proposed to enhance scene adaptability by leveraging spectral similarities among in-scene targets within local neighborhoods.
(4) Three datasets of multi-temporal HSI pairs were collected from AVIRIS data to validate the detection performance under spectral variability. Comparison experiments with classical and DL detectors validate that the proposed detector realizes a superior balance between target detectability and background suppression under spectral variability.

The remainder of this article is organized as follows. Section 2 introduces the proposed implicit contrastive learning detector (ICLTD). Section 3 presents extensive experimental results from the multi-temporal HTD datasets. Conclusions are presented in Section 4.

2. Method

2.1. Spectral Variability of Target Spectra in Multi-Temporal Images

Due to varying atmospheric, illumination, and environmental conditions, the spectral signatures of a target captured at different times may exhibit variability, which is commonly referred to as spectral variability [52]. In addition, intra-class differences in spectral characteristics may exist within the same category of targets, also resulting in variations [4]. Therefore, the prior and in-scene target spectra collected from multi-temporal HSIs may vary.

An urban-scene multi-temporal HSI pair was collected from AVIRIS data to illustrate spectral variability, as exhibited in Figure 2. These two images were captured at different periods within Jefferson County, Washington, United States. Image 1 was acquired under conditions of minimal cloud coverage, whereas Image 2 was acquired under conditions with significant cloud presence. Storage tanks were selected as the targets of interest for this example. Manual target annotations are shown in Figure 1, which were obtained with the help of ENVI software (Version 5.1). Based on these manual annotations, we computed the average target spectra from one of the images to obtain prior spectra. A comparison was then made with the target spectra in the other HSI. It is important to note that these spectra are in the radiance domain and have been normalized for better comparison.

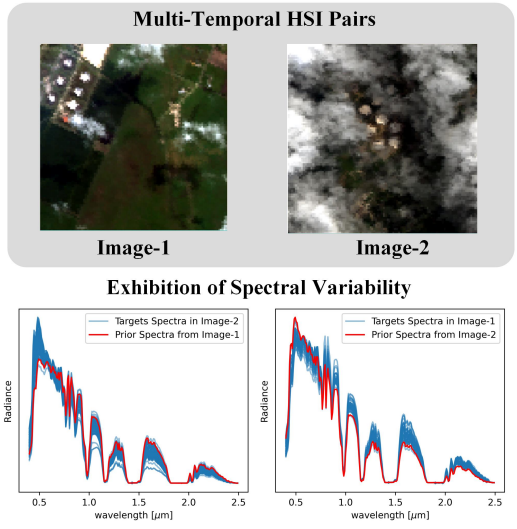

Figure 2. Exhibition of spectral variability in multi-temporal HSIs. The first row shows a pseudo-color visualization of HSIs where storage tanks were chosen as the target of interest. The prior spectra are obtained by averaging the target spectra in one of the images and are compared with the target spectra in the other HSI. The red lines in the bottom plots represent the prior spectra, while the blue lines represent the in-scene target spectra.

Based on the visualization depicted in Figure 2, notable differences can be observed between the prior spectra and in-scene target spectra. Some conventional HTD methods,

such as ACE, generally assume that the prior spectrum represents the mean of the target spectral distribution. However, in real-world scenarios, this assumption may not always hold true, thereby affecting the effectiveness of these methods. Due to the availability of only prior and in-scene spectra (without labels), certain DL-based methods utilize non-learning detectors to pre-detect background spectra. These methods generate pseudo-target spectra by combining prior spectra with the detected background spectra. The augmented data are considered pseudo-data because their labels are not manually annotated and they are used to optimize the parameters of neural networks. Simulating high-quality variational target spectra for training DL detectors is challenging because the variations in target spectra are not solely caused by spectral mixture.

2.2. Exploiting In-Scene Spectra with Implicit Contrastive Learning

The motivation of this paper is to make the detector adapt to spectral variability by exploiting in-scene spectra. Based on the exhibition in Section 2.1, it is more beneficial to explore in-scene target spectra than the pseudo-data. Because the labels of in-scene spectra are unknown, supervised contrastive learning, which teaches the detector to distinguish spectra based on annotations, cannot be used for optimization. Instead, we introduce implicit contrastive learning to exploit the in-scene spectra. Unlike original contrastive learning, which provides loss functions that need to be minimized, implicit contrastive learning does not provide any loss function and is more like a regularization process. A comparison schematic diagram is shown in Figure 3 to exhibit their differences.

For the proposed detector, an objective function is used to detect prior spectra. However, such optimization may easily collapse without regularization, i.e., when considering all the in-scene spectra as targets. One interpretation of model collapse is that the detector only learns a simple and non-discriminative representation of the prior spectra. In contrast, the introduced implicit contrastive learning aims to regularize the optimization by learning representations of the prior and in-scene spectra. This regularization helps prevent the model from over-fitting to the prior spectrum and allows it to recognize in-scene target spectra under spectral variability. To realize this, the implicit contrastive learning module (ICLM) is designed to create a gradient back-propagation channel through which the gradients of the prior spectral features (computed based on the loss function) are allocated to the in-scene spectral features. The allocation of gradients is based on the feature differences between prior spectra and in-scene spectra. As the in-scene target and background spectra inherently differ, the gradients allocated to the different in-scene spectral features vary. A specific description of this will be provided in the following subsections.

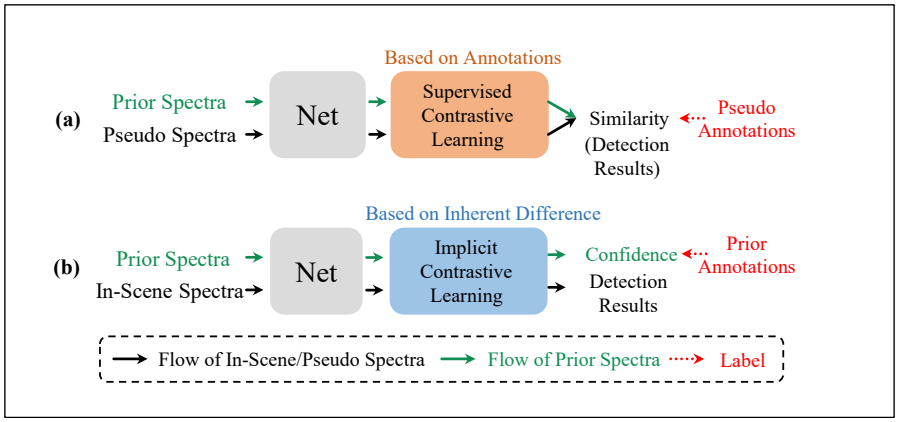

Figure 3. A comparison between (**a**) contrastive learning-based methods and (**b**) the proposed implicit contrastive learning-based detector. Contrastive learning requires the supervision of numerous annotations while implicit contrastive learning is free of annotations and plays the role of regularization. Only prior spectra provide supervision for the proposed detector, and implicit contrastive learning regularizes the detector based on the inherent differences between prior and in-scene spectra.

2.3. Structure and Pipeline of the Proposed Detector

This subsection introduces the structure and pipeline of the proposed ICLTD. An in-scene HSI is denoted $\boldsymbol{S} \in \mathbb{R}^{wh \times l}$, where w, h, and l are the width, height, and band number of the HSI, respectively. The ith in-scene spectrum of $\boldsymbol{S}$ is denoted $\boldsymbol{s}_i$. A prior spectrum captured from another HSI is denoted $\boldsymbol{s}_p \in \mathbb{R}^l$. Different from box-level target detection in the remote sensing field, existing HTD detectors commonly access in-scene spectra (without annotations) for optimization. After training, the detector outputs the target confidence scores of the in-scene spectra as detection results, which are denoted $\boldsymbol{C} \in \mathbb{R}^{wh \times 1}$. In the training phase, all of the in-scene spectra and the prior spectrum make up the data batch. For convenience, the number of in-scene spectra is denoted as $n_1 = w \times h$. In this paper, only one prior spectrum is generated from the HSI captured at different times. Hence, the batch size is $n_1 + 1$.

The network structure is shown in Figure 4. As the illumination intensity varies over time and space, significant differences in the radiance amplitude may occur. To mitigate the impact of varying amplitudes, a linear transformation process is used to normalize the Euclidean norm (L2 norm) of the in-scene spectra and prior spectra to 1. The normalized spectra are then inputted into multiple fully connected blocks (FCBs) for feature extraction. Each FCB consists of a fully connected layer, the ICLM, and a nonlinear activation layer. The FC is related to the normalized vector $\boldsymbol{s}_i$ by:

$$\boldsymbol{x}_i = \boldsymbol{W} \times \boldsymbol{s}_i + \boldsymbol{b}, \tag{1}$$

where $\times$ represents matrix multiplication. $\boldsymbol{W} \in \mathbb{R}^{l \times d}$ and $\boldsymbol{b} \in \mathbb{R}^d$ are the weight and bias of the fully connected layer, respectively, where d is the dimension number of feature representations. The FC transforms each input vector independently; that is, there is no gradient path from $\boldsymbol{x}_i$ to $\boldsymbol{s}_j$ when $i \neq j$.

The realization of implicit contrastive learning is based on the designed ICLM, which is a refinement of the original batch normalization (BN) [53]. The original BN was proposed to accelerate the training process while the ICLM is designed to regularize the optimization. The motivation for creating the ICLM is to force the detector to learn a representation of the in-scene spectra during the optimization. Without the ICLM, the detector only learns to transform the prior spectra for classification, and it cannot distinguish other spectra. The original BN could actually mitigate the above model collapse by leaking gradients from the features of the prior spectra to the in-scene spectra. However, the original BN was not specially designed for HTD and cannot provide adequate regularization because BN preserves most gradients for the learning of prior spectra rather than delegating to the in-scene spectra. Further theoretical analysis of this process is presented in Section 2.4.

To solve this problem, the ICLM duplicates the prior spectra representations for normalization to provide adequate regularization. The pipeline of the ICLM is exhibited in Figure 5. Specifically, the ICLM augments the prior spectra representation, $\boldsymbol{x}_p$, n_2 times and combines them with in-scene spectral features to compute the mean ($\boldsymbol{\mu} \in \mathbb{R}^d$) and variance ($\boldsymbol{\sigma}^2 \in \mathbb{R}^d$), where d is the dimension number of input spectral feature $\boldsymbol{x}$. ICLM is computed using $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ as:

$$\begin{cases} \boldsymbol{\mu} = \dfrac{1}{m}\left(n_2 \boldsymbol{x}_p + \displaystyle\sum_{i=1}^{n_1} \boldsymbol{x}_i\right) \\ \boldsymbol{\sigma}^2 = \dfrac{1}{m}\left(n_2 (\boldsymbol{x}_p - \boldsymbol{\mu})^2 + \displaystyle\sum_{i=1}^{n_1} (\boldsymbol{x}_i - \boldsymbol{\mu})^2\right), \end{cases} \tag{2}$$

where n_2 is the duplication number of prior spectra representation and m is the sum of n_1 and n_2.

The computed mean and variance vectors are used to normalize the features of in-scene spectra and the prior spectra:

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\sqrt{\boldsymbol{\sigma}^2}}, \tag{3}$$

where $\hat{\boldsymbol{x}}$ is the normalized feature vector.

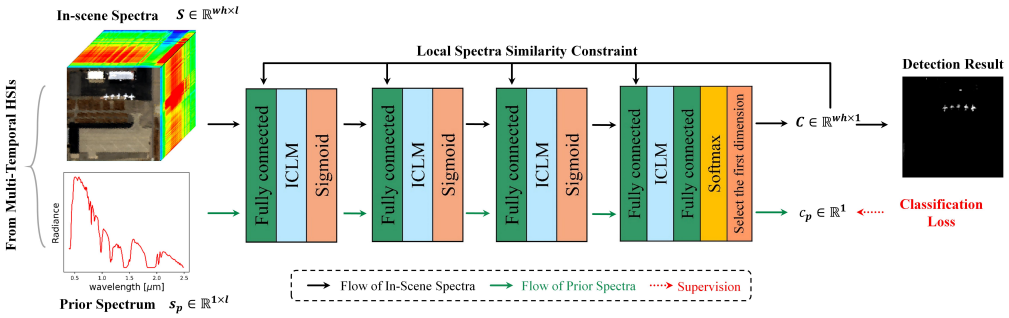

Figure 4. Network structure and training pipeline of the proposed ICLTD. The input in-scene HSI and prior spectra are captured under different imaging conditions (at different times). The ICLTD is optimized to classify the few-shot prior spectra. The ICLM is designed to regularize the optimization, which forces the detector to learn differentiated representations of the in-scene spectra. The LSSC is applied to the representations of the in-scene target spectra to improve target detectability.

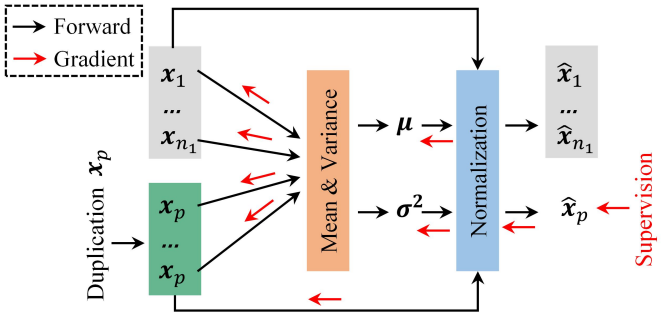

Figure 5. Pipeline of the ICLM. The black and red lines represent the flow of forward propagation and gradient back-propagation, respectively. The ICLM leaks supervised gradients from the prior spectral features ($\hat{x}_p$) to the in-scene spectral features ($\hat{x}_1, \cdots \hat{x}_{n_1}$), which forces the detector to learn representations of the in-scene spectra to minimize the loss function. Features of the prior spectra are duplicated n_2 times for adequate leaked gradient signals. The ICLM becomes the original BN when $n_2 = 1$.

Learnable vectors of the mean and variance ($\boldsymbol{\alpha}_b \in \mathbb{R}^d$ and $\boldsymbol{\beta}_b \in \mathbb{R}^d$, respectively) are used to fine-tune the distributions:

$$\boldsymbol{y} = \boldsymbol{\alpha}_b \odot \hat{\boldsymbol{x}} + \boldsymbol{\beta}_b, \tag{4}$$

where $\odot$ represents the element-wise product (Hadamard product). The complete algorithm flow of ICLM is exhibited in Algorithm 1.

Algorithm 1 Pipeline of ICLM

Input: Features of the prior spectra and in-scene HSI: $\boldsymbol{x}_p$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_{wh}$; and learnable parameters of the ICLM: $\boldsymbol{\alpha}_b$ and $\boldsymbol{\beta}_b$.
Output: Normalized spectral features: $\boldsymbol{y}_p$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{wh}$.
1: Duplicate $\boldsymbol{x}_p$ n_2 times and construct a feature batch.
2: Compute $\boldsymbol{\mu}$ and σ^2 of the feature batch following Equation (2).
3: Normalize $\boldsymbol{x}_p$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{wh}$ with $\boldsymbol{\mu}$ and σ^2 following Equation (3).
4: Fine-tune the features with $\boldsymbol{\alpha}_b$ and $\boldsymbol{\beta}_b$ following Equation (4).
5: **return** $\boldsymbol{y}_p$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{wh}$.

After using the ICLM, a sigmoid function is applied to features to increase the nonlinear feature extraction ability of ICLTD. Multiple FCBs are used to extract features of

the input spectra in a cascading manner. It is worth noting that the last FCB is free of the sigmoid layer, which brings a larger numerical range to the following classifier.

The classifier is composed of a fully connected layer and a softmax layer. The fully connected layer outputs predicted confidence vectors. The softmax layer normalizes the output confidence vectors to represent probabilities. The dimensions of the predicted probability vectors are the same as those of the latent features. The values of different dimensions represent the confidence of an in-scene spectrum belonging to different categories (targets of interest or various backgrounds). Because only the target prior is known and target spectra are our interest, the probabilities belonging to the target are selected from the output. The first dimension is assumed to be the target class by default, which is used for computing the loss function. The other dimensions of the output vectors represent the probabilities of class-agnostic backgrounds. For the visualization shown in Figure 4, the predicted target probabilities of the input images are denoted $\boldsymbol{C} \in \mathbb{R}^{wh \times 1}$. The predicted confidence of the prior spectrum and the ith in-scene spectrum of $\boldsymbol{S}$ are denoted c_p and c_i, respectively.

Unlike existing contrastive learning-based detectors, which require both positive and negative sample pairs for their loss functions, the proposed detector only utilizes the manual annotations of the prior spectrum for explicit supervised learning. The specific loss function for classifying positive samples is as follows:

$$\ell = -log(c_p). \tag{5}$$

When the ICLM is not present in the detector, the feature extraction and classification of a batch of spectra become independent of each other. This implies that the learning of detecting prior spectra does not help the detector to distinguish in-scene spectra. If optimizing the detector with Equation (5) in the absence of the ICLM, the detector will over-fit when detecting prior spectra. When the ICLM is present, the detector will learn to minimize Equation (5) by changing the representation of in-scene spectra, which regularizes the optimization.

2.4. Analysis of the ICLM

Under the HTD task, this subsection illustrates how the ICLM realizes implicit contrastive learning and brings discriminative feature extraction capability to the detector.

From the data distribution perspective, the ICLM normalizes features to restrict extreme output aggregation or divergence [51]. Therefore, the detector will not output identical results (feature aggregation) or will only be able to detect prior spectra (feature divergence). The ICLM, to some extent, mitigates the differences between the prior and target spectra by adjusting the data distributions. However, adjusting the feature distribution cannot help the optimization much. To validate the effectiveness of distribution adjustment, an ablation study is conducted in Section 3.5 where the gradient propagation component of the ICLM is disabled.

Compared to feature distribution adjustment, the gradient propagation path established by the ICLM between the in-scene spectra and prior spectra is more important. The last ICLM layer is chosen to analyze the implicit contrastive learning because the gradient calculation in this layer is simpler than those in the previous layers. Note that all of the vector multiplication presented in this subsection is the Hadamard product, which is omitted for brevity.

For convenience, the hidden features before and after the last ICLM layer are denoted $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, respectively. We define a spectral feature sequence to exhibit the gradient propagation path from $\boldsymbol{x}_p$ to unlabeled $\boldsymbol{x}$ as:

$$\boldsymbol{x}_i = \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n_1+n_2}, \tag{6}$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1}$ are in-scene spectral features and $\boldsymbol{x}_{n_1+1}, \ldots, \boldsymbol{x}_{n_2}$ are duplicated prior features. $\boldsymbol{x}_i$ equals $\boldsymbol{x}_p$ when $i > n_1$. The normalized spectral features outputted by the last ICLM are computed by substituting $\boldsymbol{\mu}$ and σ^2 in Equation (3) according to Equation (2):

$$\hat{\boldsymbol{x}}_p = \frac{m \sum_{i=1}^{m} \boldsymbol{x}_p - \boldsymbol{x}_i}{\sqrt{\sum_{i=1}^{m} (\sum_{j=1}^{m} (\boldsymbol{x}_i - \boldsymbol{x}_j))^2}}, \tag{7}$$

where only the normalized prior spectral features are exhibited because the in-scene spectral features are not used in the loss function.

According to the numerator in Equation (7), the prior spectral features are subtracted from each input feature when the ICLM normalizes the distribution means. Similarly, the denominator constructs computational relationships between any two spectral samples. Because of the established gradient propagation paths, representations of in-scene spectra also receive supervised signals during optimization.

The gradient of features before the last ICLM ($\boldsymbol{x}$) received from the ICLM is calculated to analyze the regularization of the implicit contrastive learning. There are three paths that could pass the gradients to $\boldsymbol{x}$: $\hat{\boldsymbol{x}}$, $\boldsymbol{\mu}$, and σ^2. Because the gradients $\frac{\partial \ell}{\partial \boldsymbol{\mu}}$ and $\frac{\partial \ell}{\partial \sigma^2}$ rely on $\frac{\partial \ell}{\partial \hat{\boldsymbol{x}}}$, the latter is calculated first.

Since in-scene spectral features are not used in classification loss, their derivatives are zero:

$$\frac{\partial \ell}{\partial \hat{\boldsymbol{x}}_i} = \boldsymbol{g}_i = \begin{cases} 0, & \hat{\boldsymbol{x}}_i \text{ is unlabeled} \\ \boldsymbol{g}_p, & \hat{\boldsymbol{x}}_i \text{ is prior,} \end{cases} \tag{8}$$

where $\frac{\partial \ell}{\partial \hat{\boldsymbol{x}}_i}$ is denoted $\boldsymbol{g}_i$ for convenience. When $i > n_1$, $\frac{\partial \ell}{\partial \hat{\boldsymbol{x}}_i}$ are the same because of duplication and are denoted $\boldsymbol{g}_p$.

Next, the derivative of ℓ with respect to $\boldsymbol{\mu}$ and σ^2 is calculated. According to Equation (3), the gradient of the loss function ℓ with respect to σ^2 is:

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{-n_2 \boldsymbol{g}_p \hat{\boldsymbol{x}}_p}{2\sigma^2}. \tag{9}$$

The gradient of the loss function ℓ with respect to $\boldsymbol{\mu}$ can be simplified as:

$$\frac{\partial \ell}{\partial \boldsymbol{\mu}} = \frac{-n_2 \boldsymbol{g}_p}{\sqrt{\sigma^2}}. \tag{10}$$

According to Equation (3) and Equation (2), the gradients of $\hat{\boldsymbol{x}}$, $\boldsymbol{\mu}$, and σ^2 with respect to $\boldsymbol{x}$ are, respectively:

$$\begin{cases} \frac{\partial \hat{\boldsymbol{x}}}{\partial \boldsymbol{x}} = \frac{1}{\sqrt{\sigma^2}} \\ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{x}} = \frac{1}{m} \\ \frac{\partial \sigma^2}{\partial \boldsymbol{x}} = \frac{2(\boldsymbol{x} - \boldsymbol{\mu})}{m}. \end{cases} \tag{11}$$

According to the above gradients, $\frac{\partial \ell}{\partial \boldsymbol{x}}$ can be expressed as:

$$\frac{\partial \ell}{\partial \boldsymbol{x}} = \frac{1}{m\sqrt{\sigma^2}} (\underbrace{m\boldsymbol{g}}_{①} - \underbrace{n_2 \boldsymbol{g}_p}_{②} - \underbrace{n_2 \hat{\boldsymbol{x}}_p \hat{\boldsymbol{x}} \boldsymbol{g}_p}_{③}), \tag{12}$$

where the gradients labeled as ①, ②, and ③ come from $\hat{\boldsymbol{x}}$, $\boldsymbol{\mu}$, and σ^2, respectively.

Based on the loss function (Equation 5), $g_i = 0$ for in-scene spectral features ($i < n_1$). According to Equation (12), the gradient of ℓ with respect to $\boldsymbol{x_i}$ is summarized as:

$$\frac{\partial \ell}{\partial \boldsymbol{x}_i} = \begin{cases} \frac{1}{m\sqrt{\sigma^2}}(-n_2 \boldsymbol{g}_p - n_2 \hat{\boldsymbol{x}}_p \hat{\boldsymbol{x}}_i \boldsymbol{g}_p), i \leq n_1 \\ \frac{1}{m\sqrt{\sigma^2}}(n_1 \boldsymbol{g}_p - n_2 \hat{\boldsymbol{x}}_p^2 \boldsymbol{g}_p), i > n_1 \end{cases} \tag{13}$$

Based on Equation (13), the gradients computed from ℓ are passed to in-scene spectra through $\boldsymbol{\mu}$ and σ^2 in the ICLM. In other words, the features of in-scene spectra are also optimized to realize the classification of prior spectra. Therefore, implicit contrastive learning of the in-scene spectra regularizes the optimization and prevents model collapse.

To analyze how the ICLM enables the model to learn the differentiated representation of in-scene spectra, we calculate the difference in gradients obtained through the ICLM for two in-scene spectra:

$$\frac{\partial \ell}{\partial \boldsymbol{x}_i} - \frac{\partial \ell}{\partial \boldsymbol{x}_j} = \frac{n_2 \hat{\boldsymbol{x}}_p \boldsymbol{g}_p (\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j)}{m\sqrt{\sigma^2}}. \tag{14}$$

Assuming the two in-scene spectra in Equation (14) are the target and background spectra, their inherent differences are automatically utilized by the ICLM to allocate differentiated gradients. We also calculate the difference in gradients obtained through σ^2 for the in-scene spectra and the prior spectra:

$$\frac{\partial \ell}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial \boldsymbol{x}_i} - \frac{\partial \ell}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial \boldsymbol{x}_p} = \frac{n_2 \hat{\boldsymbol{x}}_p \boldsymbol{g}_p (\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_p)}{m\sqrt{\sigma^2}}. \tag{15}$$

Based on Equation (15), the gradients received by in-scene target spectra (through σ^2) are more similar to those of the prior spectra than the in-scene background spectra, because of their inherent data similarities. In contrast, the gradients received by the in-scene background spectra are different from those of the prior spectra because of inherent data differences.

$$\sum \frac{\partial \ell}{\partial \boldsymbol{x}} = \frac{n_2 \boldsymbol{g}_p \hat{\boldsymbol{x}}_p \sum_{i=1}^{m} \hat{\boldsymbol{x}}_i}{m\sqrt{\sigma^2}} = 0. \tag{16}$$

According to Equation (16), the total amount of gradients passed through the ICLM from ℓ to $\boldsymbol{x}$ equals 0, which means that the signs of the transmitted gradients through the ICLM to the prior and in-scene spectral features are opposite. Therefore, the more prior spectra in the ICLM are duplicated, the stronger the supervised signal transferred to the unlabeled spectra will be. Therefore, n_2 determines the strength of regularization. The larger n_2 is, the more the model relies on learning the representations of in-scene spectra to minimize the loss function. When $n_2 = 1$, the ICLM becomes the original BN. Because $n_1 \gg n_2$, the gradients received by in-scene spectral features are close to 0 according to Equation (15). Hence, the original BN could not provide adequate regularization.

To summarize, in this section, we analyzed the ICLM from three perspectives. From the perspective of data distribution, the ICLM avoids excessive aggregation or divergence of the extracted features. Regarding forward propagation, the ICLM establishes gradient propagation paths between prior and unlabeled spectra. From the perspective of gradient back-propagation, the ICLM transfers the gradient of the loss function to in-scene spectra, and the feature differences are used to determine the transferred gradients. As the number of prior spectra increases in the ICLM (i.e., as n_2 increases), the regularization brought by implicit contrastive learning is enhanced.

2.5. Local Spectral Similarity Constraint

With the development of imaging spectroscopy technology, the spatial and spectral resolution of HSIs has improved. Targets in HSIs may occupy multiple pixels, as shown

by the example in Figure 6. For a multiple-pixel target, its 3D geometric structure and its relationship with the background vary from different locations, which cause differences in its spectral characteristics. Our aim is to minimize the feature differences of these locally connected spectra, thereby improving the performance of in-scene adaptability. During training, the predicted confidence scores of in-scene spectra can be used to select candidate targets. For each candidate sample, if there are samples with higher confidence within their 3×3 neighborhoods, we increase the similarity between the feature representations of these two samples. Note that the LSSC can be employed if a target occupies two or more connected pixels.

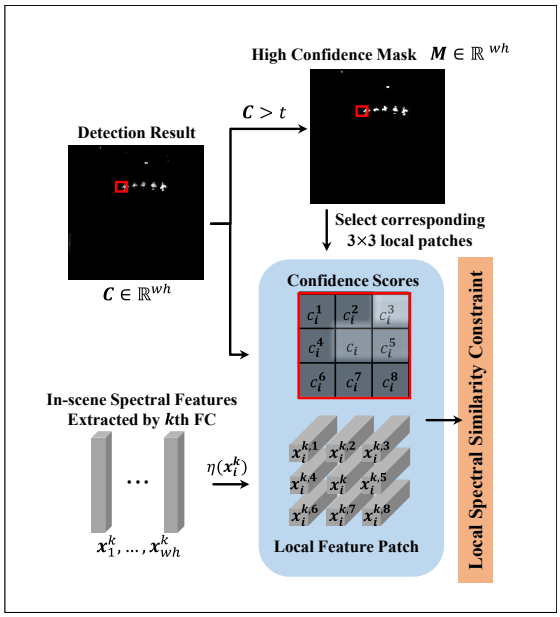

Figure 6. Process diagram of the LSSC. The LSSC is based on the assumption that the spectral samples of multi-pixel targets are similar and spatially connected. The detection results (c_i) are used to collect candidate targets ($\boldsymbol{x}_i^k$) as the centers of 3×3 local patches ($\eta(\boldsymbol{x}_i^k)$). The candidate in-scene spectra are optimized to be similar to surrounding spectra whose confidence scores are higher than that of the center. An example is shown in the figure. The detection results of the left wing of the aircraft are shown in the red box. Because $c_i < c_i^3$, the LSSC enhances the similarity between $\boldsymbol{x}_i^k$ and $\boldsymbol{x}_i^{k,3}$.

According to the above motivation, the LSSC was proposed and applied to representations of in-scene spectra outputted by each fully connected layer of the FCBs. First, we find the target neighborhoods. A threshold, t, is set to select high-confidence in-scene spectra. Given the detection results of in-scene HSI, $\boldsymbol{C}$, the mask that reflects the distribution of target candidates is:

$$\boldsymbol{M} = \boldsymbol{C} > t. \tag{17}$$

Latent features of the in-scene spectra extracted by the kth fully connected layer are denoted $\{\boldsymbol{x}_1^k, \ldots, \boldsymbol{x}_{wh}^k\}$. Their corresponding confidence scores are denoted $\{c_1, \ldots, c_{wh}\}$. According to $\boldsymbol{M}$, the representations of target candidates are collected into the set: $\mathbf{T}^k = \{\boldsymbol{x}_i^k | \boldsymbol{M}_i > 0\}$. For each candidate, its 3×3 neighborhood is constructed, encompassing multi-level representations and confidence scores. The neighborhood of features is denoted $\{\boldsymbol{x}_i^{k,1}, \boldsymbol{x}_i^{k,2}, \ldots, \boldsymbol{x}_i^{k,8}\}$, where $\boldsymbol{x}_i^{k,1}$ represents the first neighboring feature of $\boldsymbol{x}_i^k$. The corresponding confidence scores are $\{c_i^1, c_i^2, \ldots, c_i^8\}$. For each $\boldsymbol{x}_i^k$, the set consisting of its neighboring target spectral features is denoted $\boldsymbol{\eta}(\boldsymbol{x}_i^k) = \{\boldsymbol{x}_i^{k,j} | c_i < c_i^j\}$, where c_i^j represents the target confidence score of $\boldsymbol{x}_i^{k,j}$. With the above features, the proposed LSSC is defined as:

$$\ell_{LSSC} = \frac{-1}{n_s} \sum_{k=1}^{4} \sum_{\boldsymbol{x} \in \mathbf{T}^k} \sum_{\boldsymbol{x}' \in \boldsymbol{\eta}(\boldsymbol{x})} log\left(\frac{sg(f(\boldsymbol{x}')) \cdot f(\boldsymbol{x})}{||sg(f(\boldsymbol{x}'))||_2 ||f(\boldsymbol{x})||_2}\right), \tag{18}$$

where $sg(\boldsymbol{x}')$ represents stopping the gradient propagation of $\boldsymbol{x}'$, f is the softmax operation, and n_s is the number of candidate targets. Stopping the gradients of neighboring features is to ensure that the LSSC enhances the detectability of target candidates rather than diminishing that of the neighboring targets. The complete calculation process of LSSC is described in Algorithm 2. The total optimization function, L, is the combination of positive sample classification loss and the LSSC:

$$L = \ell + \ell_{LSSC}. \tag{19}$$

Algorithm 2 Pipeline for computing LSSC

Input: Spectral features extracted by the kth fully connected layer: $\{\boldsymbol{x}_1^k, \ldots, \boldsymbol{x}_{wh}^k\}, k \in \{1, 2, 3, 4\}$; confidence scores of in-scene spectra: $c_1, \ldots, c_{wh}$; and the threshold of the LSSC: t.
Output: The loss of the LSSC, ℓ_{LSSC}.

1: Generate the target candidate mask $\boldsymbol{M}$ according to predicted confidence scores following Equation (17).
2: Collect features of candidate in-scene targets: $\mathbf{T}^k = \{\boldsymbol{x}_i^k | \boldsymbol{M}_i > 0\}$.
3: Collect confidence scores of the candidates $\{c_i | \boldsymbol{M}_i > 0\}$.
4: Collect neighboring spectral features of each $\boldsymbol{x}_i^k$ in $\mathbf{T}^k$: $\{\boldsymbol{x}_i^{k,1}, \boldsymbol{x}_i^{k,2}, \ldots, \boldsymbol{x}_i^{k,8}\}$.
5: Collect confidence scores of neighboring features: $\{c_i^1, c_i^2, \ldots, c_i^8\}$.
6: Collect desired neighboring features of $\boldsymbol{x}_i^k$: $\boldsymbol{\eta}(\boldsymbol{x}_i^k) = \{\boldsymbol{x}_i^{k,j} | c_i < c_i^j\}$.
7: Get ℓ_{LSSC} following Equation (18).
8: **return** ℓ_{LSSC}.

3. Experiments

3.1. Experimental Setup

3.1.1. Datasets

We conducted experiments using three pairs of multi-temporal HSIs collected from AVIRIS data (https://aviris.jpl.nasa.gov/, accessed on 23 December 2023). AVIRIS is an airborne imaging spectrometer instrument capable of capturing 224 contiguous spectral radiance images in the wavelength range of 400–2500 nm. The AVIRIS data used in this study consist of the hyperspectral data collected during ground observation experiments when AVIRIS was mounted on an aircraft. The AVIRIS data have been ortho-corrected and underwent radiometric calibration.

Specifically, we collected three sets of multi-temporal HSIs from airport, beach, and urban scenes. These datasets are referred to as MT-ABU for convenience. The groundtruth maps of targets were annotated using ENVI software. The pseudo-color images and corresponding annotations are shown in Figure 7. Further details of the MT-ABU dataset can be found in Table 1.

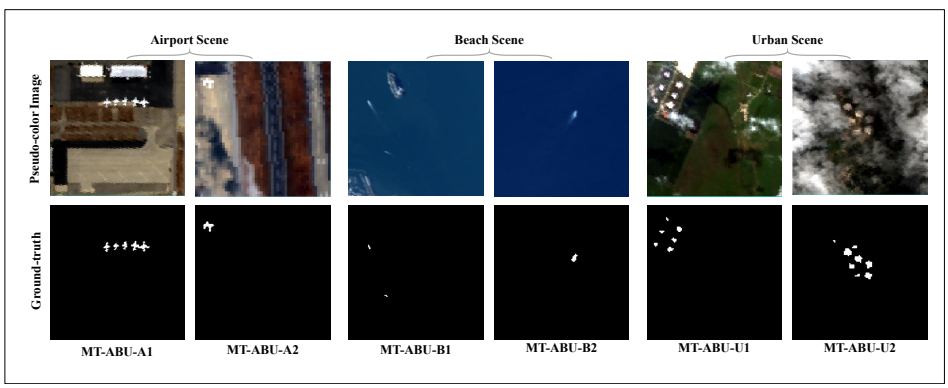

Figure 7. Pseudo-color images (first row) and ground-truth maps (second row) of six HSIs.

Each scene consists of two images captured at different times to form an image pair. Although the targets in each pair of images are not identical, they belong to the same category. The spatial resolution of each image was estimated based on the flying altitude. In cases where the AVIRIS data did not provide fly altitudes, the spatial resolution was estimated using information from Google Maps.

Table 1. Detailed features of the collected multi-temporal datasets.

Dataset	Place	Time	Spatial Resolution	Target Pixel	Image Size	Source of Prior Spectra
MT-ABU-A1	Riverside	4 August 2013	3.8 m	185	150 × 150	MT-ABU-A2
MT-ABU-A2	San Francisco Bay Area	23 June 2008	13.7 m	8	50 × 50	MT-ABU-A1
MT-ABU-B1	Long Beach	19 April 2014	16.7 m	14	150 × 150	MT-ABU-B2
MT-ABU-B2	Bay Area	22 November 2013	16.7 m	16	100 × 100	MT-ABU-B1
MT-ABU-U1	Jefferson County	29 August 2010	7.1 m	51	100 × 100	MT-ABU-U2
MT-ABU-U2	Jefferson County	28 August 2010	7.1 m	118	100 × 100	MT-ABU-U1

We conducted HTD experiments in a target-rediscovery way. Specifically, two HSIs of each scene provided prior spectra for each other. For example, the prior spectrum used for detecting targets in MT-ABU-A1 was generated by averaging the target spectral radiances from MT-ABU-A2 based on the ground-truth annotations. As a result, six images were used in the experiments, each with its corresponding prior spectrum. The in-scene HSIs and their corresponding prior spectra sources are presented in Table 1. For convenience, we will refer to each experimental dataset by the name of its respective in-scene HSI.

3.1.2. Comparison Methods

In order to validate the performance of the proposed method under the spectral variability condition, both learning and non-learning methods were employed for comparison. Three DL-based methods that required pseudo-data generation were compared: Siamese fully connected target detector (SFCTD) [32], two stream convolutional target detector (TSCNTD) [29], and Siamese transformer target detector (STTD) [31]. Non-deep learning methods included a classical method (ACE), a spectral distance-based method (R-SA^2) [12]), and the improved classical methods hierarchical CEM (HCEM) [14] and ensemble-based CEM (ECEM) [15]. Note that all of the methods utilized the same prior spectral information and did not employ any additional data; they used just the prior spectra and in-scene HSIs. Therefore, the experiments provide a fair comparison.

3.1.3. Evaluation Criteria

Three-dimensional (3D)-receiver operator characteristic (ROC) curves [54], area under the curve (AUC), and detection map visualization were used to evaluate the performances comprehensively. The training and inference times of DL methods and the processing times of the non-learning methods were used to evaluate their efficiency. The 3D-ROC curve reflects the relationship between detection probability, P_D, false alarm probability, P_F, and confidence threshold, τ. The desired curve has a lower P_F and a higher τ for the same P_D.

The AUC values of 3D-ROC curves quantitatively reflect the quality of the curves, which are denoted $\mathrm{AUC}_{(D,F)}$, $\mathrm{AUC}_{(D,\tau)}$, and $\mathrm{AUC}_{(F,\tau)}$. $\mathrm{AUC}_{(D,F)}$ measures the ability of the detector to discriminate between targets and backgrounds. $\mathrm{AUC}_{(D,\tau)}$ and $\mathrm{AUC}_{(F,\tau)}$ reflect the average confidence values of targets and backgrounds, respectively. Because $\mathrm{AUC}_{(D,\tau)}$ and $\mathrm{AUC}_{(F,\tau)}$ are unable to evaluate target–background separation ability, two derived AUC values that are proposed in this work [12] are utilized to replace them:

$$\begin{cases} \mathrm{AUC}_{TDBS} = & \mathrm{AUC}_{(D,\tau)} - \mathrm{AUC}_{(F,\tau)}, \\ \mathrm{AUC}_{SNPR} = & \mathrm{AUC}_{(D,\tau)} / \mathrm{AUC}_{(F,\tau)}. \end{cases} \tag{20}$$

AUC_{TDBS} reflects the difference in confidence scores between targets and backgrounds, while AUC_{SNPR} measures the ratio between them. Since the number of background

samples in an HSI is much larger than the number of target samples, changes in confidence scores for a subset of background samples may not significantly affect the overall amplitude but can still impact background suppression performance. In such cases, AUC_{SNPR}, which measures the differences in magnitude, provides a better reflection of performance changes. However, it is worth noting that even though some results may have high AUC_{SNPR} values, both the confidence scores for the target and the background may still be low. In such situations, AUC_{TDBS} is a more suitable metric for evaluation.

3.1.4. Implementation Details

The output dimension of each fully connected layer, d, was 50, including feature extraction and the final classifier. For convenience, the ratio of prior spectral features to in-scene spectral features in the ICLM is denoted $r = n_2/n_1$. Unless otherwise specified, r was set to 0.5 and the confidence threshold t in the LSSC was set to 0.3. These two parameters were fixed for all the comparison experiments. Parameter sensitivity experiments of r and t were conducted. During training, all of the in-scene spectra were put into a data batch. The Adam optimizer was used to optimize the model parameters, with a learning rate of 1×10^{-4} and a weight decay of 5×10^{-4}. The number of training epochs was set to 500.

The configuration of the server used for the experiment consisted of an Intel i9-9900x CPU, NVIDIA TITAN Xp, and 32 GB of RAM. The operating system was Windows 11. The Pytorch package was used for running the DL-based approach codes. When comparing the detection performance, all methods used the same prior spectra. Non-learning methods were run using the NumPy package in the Python environment. For all the methods, the L2-norm of in-scene and prior spectra were normalized to 1 for better performance.

3.2. Parameter Sensitivity Experiments

The proposed HTD method consists of two main components: the ICLM and LSSC. Both modules have parameters that are manually set, r and t. r is the ratio of the number of prior spectra to in-scene spectral features in the ICLM, and t is a confidence threshold for collecting target candidates. In order to analyze the influence of these two parameters on performance, we conducted experiments with different parameter settings. The tested values for r were $\{0.1, 0.25, 0.5, 0.75, 1\}$ and for t were $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. In total, 25 experiments were performed on each dataset, and the AUC results of each experiment are displayed in a 3D bar graph in Figure 8.

The $AUC_{(D,F)}$ results under different t are stable and excellent for the six datasets. Although there is a slight fluctuation in the $AUC_{(D,F)}$ results for the MT-ABU-A1 data, they remain at a high level. The differences in the AUC_{TDBS} and AUC_{SNPR} results under the different settings of t are not significant. According to the above results, the ICLTD has relatively weak sensitivity to the variable t. In practical applications, t has the flexibility to assume values across a wide range without significantly influencing the performance.

The $AUC_{(D,F)}$ results with different r are also stable and great. However, the AUC_{TDBS} and AUC_{SNPR} results significantly changes for different values of r. As r increases, AUC_{TDBS} decreases while AUC_{SNPR} increases. Although the changes are significant, the AUC_{TDBS} and AUC_{SNPR} results are promising. According to the illustration given in Section 2.4, r determines the regularization strength brought by implicit comparison learning. The above experimental results reflect that an increase in regularization strength improves the background suppression ability while impacting the detector's ability to detect variational in-scene target spectra. In practical applications, if spectral variability is strong, a small r could ensure target detectability. On the other hand, when spectral variability is not obvious, a larger r could be set to achieve better background suppression performance.

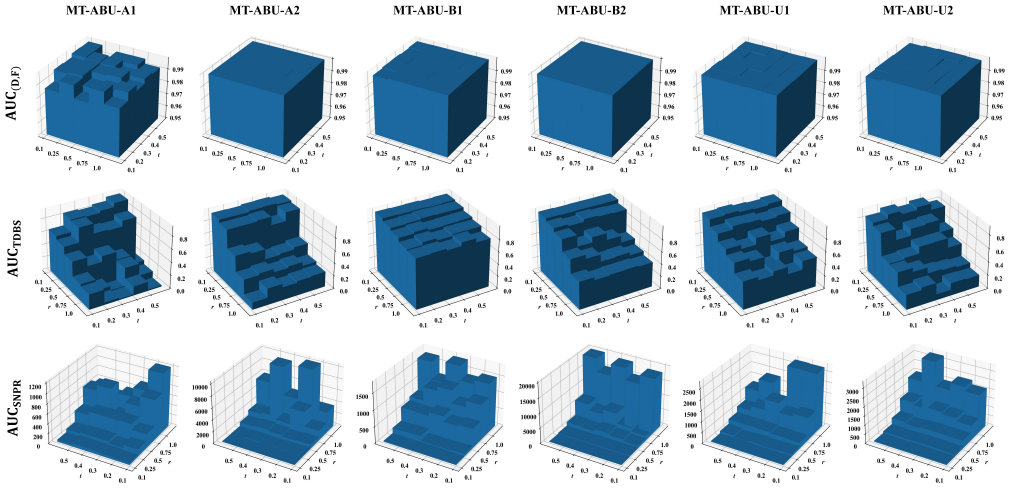

Figure 8. Visualization of the AUC results ($AUC_{(D,F)}$, AUC_{TDBS}, and AUC_{SNPR}) of the ICLTD for different parameters (r and t). r is the ratio of the number of prior spectra to in-scene spectral features in the ICLM. t is a confidence threshold for collecting target candidates.

3.3. Visualization of Feature Distance during Training

According to the theoretical analysis illustrated in Section 2.4, implicit contrastive learning of in-scene spectra regularizes the optimization and allocates various gradients for discriminative representations of in-scene spectra. Moreover, the LSSC was designed to reduce variance among the in-scene target spectra representations. To study how the ICLM and LSSC influence feature extraction, we computed the feature difference between in-scene spectra and prior spectra during the training process. The L1 distance was applied to features extracted by the last FCB. For convenience, the distance between in-scene targets and priors is denoted $d_{(t,p)-\mathrm{LSSC}}$, and the distance between in-scene backgrounds and priors is denoted $d_{(b,p)-\mathrm{LSSC}}$. The above two distances are denoted $d_{(t,p)}$ and $d_{(b,p)}$, respectively, if the LSSC was not applied to the optimization. The distances of all the targets and backgrounds were averaged for visualization, as exhibited in Figure 9.

Because the parameters of the ICLTD were randomly initialized, $d_{(t,p)}$ and $d_{(b,p)}$ were small at the beginning. During the optimization process, $d_{(b,p)}$ consistently increased, and $d_{(t,p)}$ also increased during the first few epochs. Then, $d_{(t,p)}$ started to decline and eventually stabilized or increased slowly. Because of the different growth paths of $d_{(b,p)}$ and $d_{(t,p)}$, the optimized detector learns differentiated representations of in-scene spectra. These results validate that the introduced implicit contrastive learning of in-scene spectra prevents model collapse and can help the ICLTD distinguish in-scene target spectra from background spectra. Because the regularization of implicit contrastive learning is based on inherent data differences, some variational in-scene targets are also suppressed due to spectral variability, which is the reason why $d_{(b,p)}$ may increase slowly near the end of the process.

When LSSC was applied to the optimization, $d_{(b,p)-\mathrm{LSSC}}$ also consistently increased, eventually equaling $d_{(b,p)}$, which validates why the LSSC is automatically applied to the in-scene targets and does not have a significant impact on background suppression. The trends of $d_{(t,p)}$ and $d_{(t,p)-\mathrm{LSSC}}$ exhibited similarities in the initial epochs. However, as training progressed, $d_{(t,p)-\mathrm{LSSC}}$ consistently showed a notably lower value compared to $d_{(t,p)}$. Furthermore, $d_{(t,p)-\mathrm{LSSC}}$ either stabilized or continued to decrease in the later stages of training. These findings provide evidence that the introduction of the LSSC effectively reduces feature variations between the prior and in-scene spectra. Consequently, it mitigates the impact caused by spectral variability.

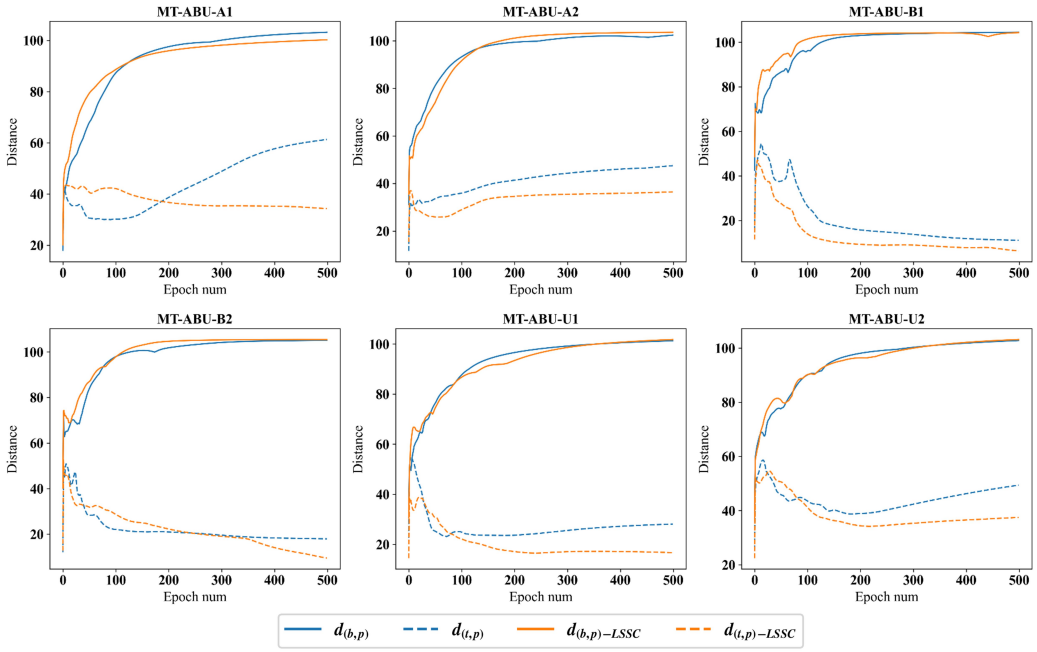

Figure 9. L_1 feature distance between the features of the prior spectra and in-scene target/background spectra during training. $d_{(t,p)-\mathrm{LSSC}}$ and $d_{(b,p)-\mathrm{LSSC}}$ represent the target-prior and background-prior distances, respectively. The above distances are denoted $d_{(t,p)}$ and $d_{(b,p)}$ for cases when the LSSC was not applied.

3.4. Comparison with Other Approaches

3.4.1. 3D-ROC Curves

We first analyzed the results of $\mathrm{ROC}_{(P_D,P_F)}$, shown in the third left column of Figure 10. It is desirable for P_D to be higher for a given P_F. For all datasets except MT-ABU-B2, the proposed ICLTD has the best $\mathrm{ROC}_{(P_D,P_F)}$ results. For the MT-ABU-B2 dataset, the $\mathrm{ROC}_{(P_D,P_F)}$ of the ICLTD is competitive. Among the three DL methods that were compared, SFCTD has the best $\mathrm{ROC}_{(P_D,P_F)}$ results, with STTD coming in second place. The $\mathrm{ROC}_{(P_D,P_F)}$ results of HCEM are better and more stable than for ECEM. For all of the HSIs except MT-ABU-B2, the $\mathrm{ROC}_{(P_D,P_F)}$ results of ACE and R-SA^2 are inferior compared to those of STTD, STTD, HCEM, and ECEM. The $\mathrm{ROC}_{(P_D,P_F)}$ results of TSCNTD are inferior to the other methods for nearly all datasets.

The experimental results of $\mathrm{ROC}_{(P_D,\tau)}$ and $\mathrm{ROC}_{(P_F,\tau)}$ were analyzed together. These two curves reflect the target detectability and background suppression of the detector, respectively, and the difference between them reflects target–background separability. The desired $\mathrm{ROC}_{(P_D,\tau)}$ has a higher P_D at given τ, while the desired $\mathrm{ROC}_{(P_F,\tau)}$ has the opposite behavior. The proposed ICLTD, STTD, and HCEM have desirable $\mathrm{ROC}_{(P_D,\tau)}$ and $\mathrm{ROC}_{(P_F,\tau)}$ results, while the ICLTD makes the best balance between these two curves. For ECEM and SFCTD, they have great $\mathrm{ROC}_{(P_D,\tau)}$ results but unsatisfactory $\mathrm{ROC}_{(P_F,\tau)}$ results. For ACE and R-SA^2, their $\mathrm{ROC}_{(P_F,\tau)}$ results are desirable but their $\mathrm{ROC}_{(P_D,\tau)}$ results are not ideal.

The 3D-ROC combines the above three 2D-ROC curves together. For convenience, we defined a coordinate order (τ, P_F, P_D) and positioned two planes in 3D coordinate space, where the green one contains samples with $\tau = 1$ while the gray one contains samples with $\tau = 0$. As shown in the visualization in Figure 10, each 3D-ROC is from (1, 0, 0) to (0, 1, 1). The ideal 3D-ROC should first increase P_D and then P_F. The curve should be close to the green plane when ascending, which reflects great target detectability. Meanwhile, the desired curve should be close to the gray plane when the value of P_F increases, proving its background suppression ability. For instance, the best curve will connect the following coordinates with a straight line: (1, 0, 0), (1, 0, 1), (0, 0, 1), and (0, 1, 1). Based on the results, the curves of the ICLTD are superior to those of the other methods in terms of balancing

target detectability and background suppression. The curves of the ICLTD first ascend then maintain high confidence scores. When the P_F increases, the curve of the ICLTD is close to the gray plane. Although the curves of ECEM are closer to the green plane when ascending, they are far from the gray plane when P_F increases. The curves of ACE and R-SA^2 are far from the green plane when ascending.

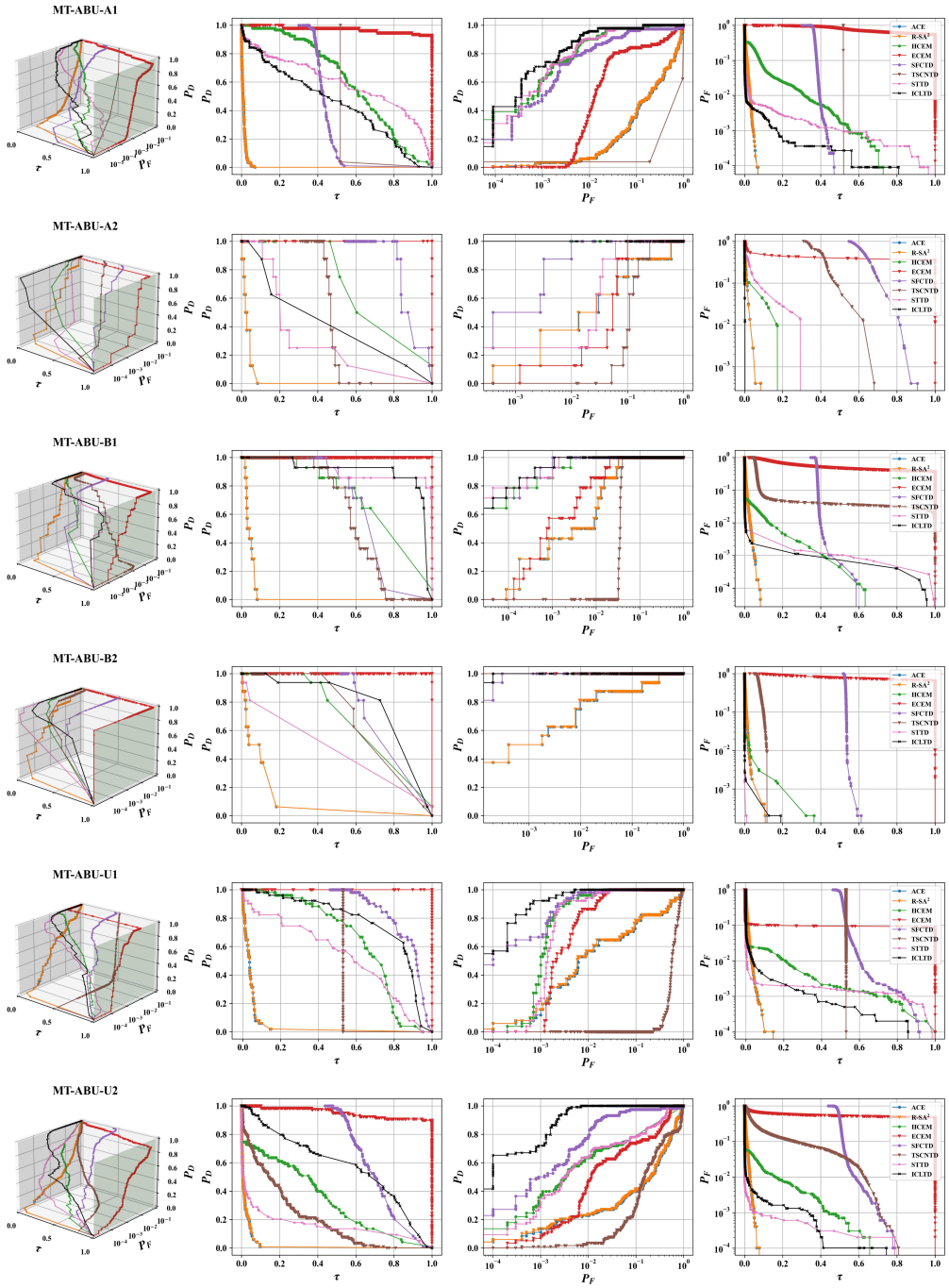

Figure 10. 3D-ROC curves results for the six MT-ABU datasets. The black lines represent the curves of the proposed ICLTD.

In summary, the proposed ICLTD has the best overall detection performance, which balances the performance of distinguishing targets and backgrounds ($\mathrm{ROC}_{(P_D,P_F)}$), target detectability ($\mathrm{ROC}_{(P_D,\tau)}$), and background suppression ($\mathrm{ROC}_{(P_F,\tau)}$). HCEM has the second-best detection performance, followed by SFCTD and STTD. SFCTD is better than STTD in

terms of distinguishing targets and backgrounds while STTD has better target–background separability. ECEM is not as good as the above methods but is still better than TSCNTD, ACE, and R-SA^2. TSCNTD, ACE, and R-SA^2 do not perform well on multi-temporal HTD datasets. The unsatisfactory performance of TSCNTD may come from optimization difficulties resulting from the absence of normalization layers. The inferior performances of ACE and R-SA^2 may result from the variations in the prior spectra.

3.4.2. AUC Values

AUC values can quantitatively reflect the quality of ROC curves. The $AUC_{(D,F)}$ results are shown in Table 2. The proposed method has the highest $AUC_{(D,F)}$ results for all datasets. HCEM, SFCTD, and STTD also achieved promising $AUC_{(D,F)}$ results, followed by ECEM, ACE, and R-SA^2. TSCNTD did not perform well in terms of $AUC_{(D,F)}$.

Table 2. $AUC_{(D,F)}$ results for the MT-ABU datasets. The best results are indicated in bold.

Method \ Dataset	A1	A2	B1	B2	U1	U2
ACE	0.6785	0.8974	0.9909	0.9661	0.8690	0.7125
R-SA^2	0.6805	0.8996	0.9912	0.9663	0.8701	0.7147
HCEM	0.9934	**1.0000**	0.9997	**1.0000**	0.9980	0.8612
ECEM	0.9048	0.9501	0.9958	**1.0000**	0.9950	0.8880
SFCTD	0.9697	0.9975	**0.9999**	**1.0000**	0.9987	0.9754
TSCNTD	0.2921	0.8768	0.9647	**1.0000**	0.3980	0.7060
STTD	0.9939	0.9743	0.9998	**1.0000**	0.9974	0.9263
ICLTD	**0.9956**	**1.0000**	**0.9999**	**1.0000**	**0.9997**	**0.9993**

The AUC_{TDBS} results are shown in Table 3. For all datasets, the proposed ICLTD maintained a competitive AUC_{TDBS} performance. HCEM also showed a stable and great AUC_{TDBS} performance. The AUC_{TDBS} results of STTD are the best for the MT-ABU-A1 and MT-ABU-B1 datasets, but are not ideal for the MT-ABU-A2 and MT-ABU-U2 datasets. Similarly, the AUC_{TDBS} results of ECEM are competitive but not promising for the MT-ABU-A1 and MT-ABU-B2 datasets. SFCTD is inferior to ICLTD, HCEM, STTD, and ECEM in terms of AUC_{TDBS}, but superior to ACE, R-SA^2, and TSCNTD.

The AUC_{SNPR} results are shown in Table 4. The proposed ICLTD shows the highest overall AUC_{SNPR} results, followed by STTD. HCEM is the third-best method in terms of AUC_{SNPR} results, followed by ACE, R-SA^2, SFCTD, and ECEM. TSCNTD performed the worst among all methods.

According to the above results, the confidence differences between targets and backgrounds of the proposed method were significant for all the datasets from both the amplitude and ratio perspectives. HCEM also performed well from the amplitude perspective but not as well as the ICLTD in terms of amplitude ratio. The target–background separation performances of the other approaches were not as stable as those of HCEM and the ICLTD.

Table 3. AUC_{TDBS} results for the MT-ABU datasets. The best results are indicated in bold.

Method \ Dataset	A1	A2	B1	B2	U1	U2
ACE	0.0064	0.0247	0.0391	0.1023	0.0413	0.0178
R-SA^2	0.0065	0.0248	0.0393	0.1024	0.0417	0.0179
HCEM	0.5844	**0.6767**	0.7185	0.6808	0.6391	0.3264
ECEM	0.1701	0.5862	0.4510	0.1973	**0.9081**	0.4017
SFCTD	0.0695	0.2473	0.2767	0.2299	0.3570	0.2013
TSCNTD	−0.0371	0.0665	0.5022	0.6141	0.0000	0.1172
STTD	**0.5906**	0.2736	**0.9070**	0.4578	0.5235	0.1543
ICLTD	0.4417	0.4125	0.8994	**0.7838**	0.7698	**0.5748**

Table 4. AUC_{SNPR} results for the MT-ABU datasets. The best results are indicated in bold.

Method \ Dataset	A1	A2	B1	B2	U1	U2
ACE	2.449	6.319	10.400	29.094	10.561	5.062
R-SA^2	2.474	6.376	10.498	29.267	10.724	5.127
HCEM	21.997	61.046	149.261	591.155	112.825	55.556
ECEM	1.213	2.420	1.852	1.254	10.222	1.731
SFCTD	2.036	3.723	9.439	19.885	7.662	4.285
TSCNTD	0.197	1.715	9.674	20.905	0.606	2.420
STTD	231.164	10.794	538.713	**13,189.051**	320.844	343.675
ICLTD	**409.669**	**789.488**	**717.131**	2158.172	**348.794**	**471.063**

3.4.3. Detection Map Visualization

Detection map visualization can intuitively reflect the saliency of targets and the contrast between targets and backgrounds. Here, to show real contrast, the original detection results of HTD methods (except HCEM and ECEM) were visualized without normalization. When the maximum values outputted by HCEM and ECEM were larger than 1, their detection maps were normalized to 0–1 for visualization.

The visualization results are shown in Figure 11. The ICLTD is highly effective in providing accurate target distribution while suppressing the background spectra, resulting in optimal target–background contrast. The visualization results of HCEM are the second-best, whose background suppression performance was slightly worse than that of the ICLTD. However, its achieved target saliency for the MT-ABU-A2 dataset was even better than that achieved with the ICLTD. STTD and SFCTD both provided fairly comprehensive target distributions. STTD showed excellent target–background contrast but more false alarms. SFCTD presented a slightly lower target–background contrast compared to STTD. ECEM, ACE, and R-SA showed reasonable performance for certain datasets. However, they performed poorly for specific datasets, such as the MT-ABU-A1 dataset. According to the detection map visualization, TSCNTD only performed well for the MT-ABU-B2 dataset.

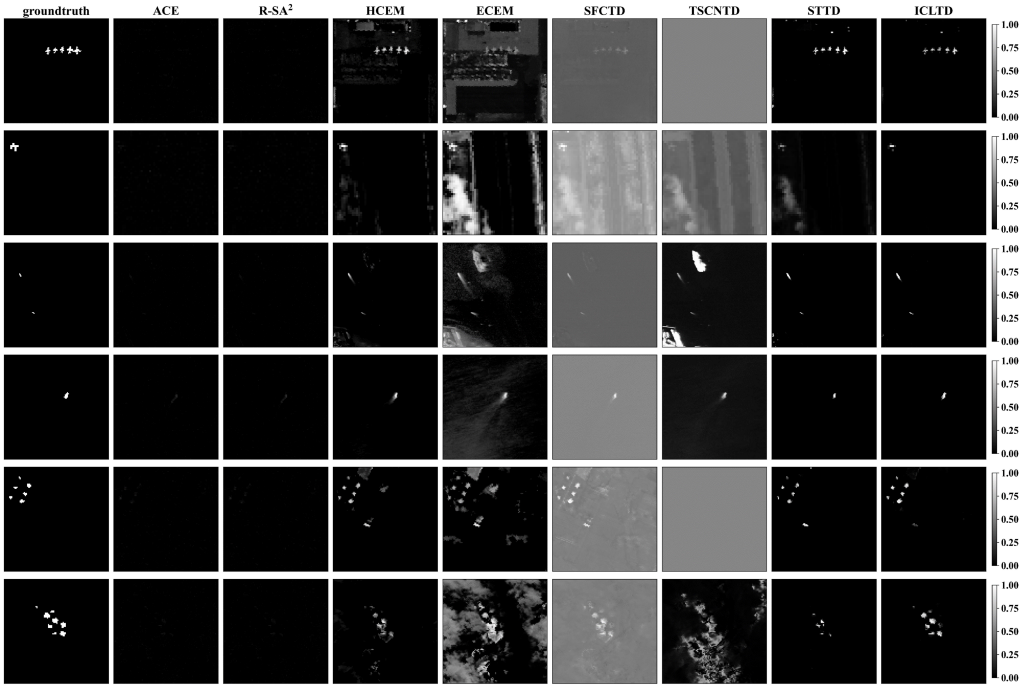

Figure 11. Detection map comparison for the MT-ABU datasets.

3.4.4. Inference Time Comparison

The training times of the DL methods were used for comparison because these methods need to retrain their respective parameters for new in-scene HSIs. However, when there is a large amount of data, DL-based methods can train models before detecting in-scene HSIs, just like object detection in RGB images [55]. In that case, these methods only require a small amount of time (inference time) to complete the detection task.

The consumed times of the ICLTD and the other methods are exhibited in Table 5. We have also exhibited the inference times of the DL methods for reference. Of the compared methods, ACE and R-SA^2 are the most efficient. HCEM and ECEM take less time than the DL methods. In terms of DL methods, SFCTD and the ICLTD are more efficient due to their composition of fully connected layers. On the other hand, TSCNTD consists of multiple 1D convolution layers, and STTD includes several self-attention modules, resulting in longer training times compared to SFCTD and the ICLTD. Although the training times of the DL methods are longer than the time consumed by the classical approaches, their inference times are very fast. If the trained models can be directly applied to other similar in-scene HSIs, the efficiency of DL methods will be much higher than non-learning methods.

Table 5. Time consumption (s) of traditional approaches and DL methods on six datasets. * represents the inference time of the DL methods. The best results are shown in bold.

Method \ Dataset	A1	A2	B1	B2	U1	U2
ACE	**0.784**	**0.551**	**0.574**	0.752	0.758	**0.542**
R-SA^2	0.872	0.723	0.694	**0.475**	**0.392**	0.650
HCEM	14.942	1.374	2.022	5.916	2.734	3.445
ECEM	13.532	3.666	7.418	8.186	14.156	7.778
SFCTD	50.964	4.215	15.513	16.247	47.035	15.475
TSCNTD	90.925	10.716	35.577	37.047	89.884	36.882
STTD	133.413	14.486	53.971	54.563	133.266	54.133
ICLTD	37.193	14.245	21.611	21.457	35.574	22.234
SFCTD *	0.019	**0.003**	**0.007**	**0.007**	**0.014**	**0.004**
TSCNTD *	0.169	0.052	0.066	0.063	0.145	0.062
STTD *	3.061	0.281	1.376	1.188	2.883	1.357
ICLTD *	**0.018**	0.015	**0.007**	0.008	0.030	0.015

3.5. Ablation Study

The results in this section are used to validate and discuss the effectiveness of the ICLM and LSSC. The ICLM realizes implicit contrastive learning using in-scene spectra. It aims to regularize the learning process of classifying few-shot prior spectra by enabling the model to learn differentiated representations of in-scene spectra, rather than over-fitting the few-shot prior spectra. According to the analysis in Section 2.4, although the ICLM changes the input feature distributions, the work of the ICLM relies on sufficient gradient signals that propagate between the prior spectra and in-scene spectra.

In order to validate the aforementioned analysis and the effectiveness of the LSSC, the following four experiments were conducted:

1. ICLM removal.
2. Zeroing gradient signals in the ICLM.
3. Replacing the ICLM with the original batch normalization (BN).
4. LSSC removal.

Experiments 1–3 were conducted with the LSSC removed. The first experiment aimed to validate the effectiveness of the ICLM; the second aimed to validate that the effectiveness of the ICLM lies in signal transmission rather than changes in feature distribution; and the third experiment was designed to study the consequences of insufficient gradient signal transmission. Experiment 4 was conducted to validate the effectiveness of the

LSSC. The results of these four experiments are denoted ICLTD-w/o-ICLM, ICLTD-ZG, ICLTD-BN, and ICLTD-w/o-LSSC, respectively.

The visualization results are exhibited in Figure 12. The AUC results are presented in Tables 6–8. After removing the ICLM, optimization of the ICLTD collapsed and all the in-scene spectra were considered as targets with higher confidence. The AUC results and visualization of the ICLTD-w/o-ICLM are poor. The results of Experiment 1 prove the effectiveness of the ICLM in regularizing the optimization.

Table 6. $AUC_{(D,F)}$ results for the ablation studies using the MT-ABU datasets. The best results are indicated in bold.

Exp \ Dataset	A1	A2	B1	B2	U1	U2
ICLTD-w/o-ICLM	0.8983	0.7621	0.9895	**1.0000**	0.9713	0.4907
ICLTD-ZG	0.7931	0.9891	0.9993	0.9994	0.8453	0.6297
ICLTD-BN	0.9949	**1.0000**	0.9995	0.9985	0.9867	0.9235
ICLTD-w/o-LSSC	0.9950	**1.0000**	**0.9999**	**1.0000**	0.9989	0.9991
ICLTD	**0.9956**	**1.0000**	**0.9999**	**1.0000**	**0.9997**	**0.9993**

Table 7. AUC_{TDBS} results for the ablation studies using the MT-ABU datasets. The best results are indicated in bold.

Exp \ Dataset	A1	A2	B1	B2	U1	U2
ICLTD-w/o-ICLM	0.0000	0.0053	0.0003	0.0011	0.0002	0.0001
ICLTD-ZG	0.2407	0.8372	0.9084	**0.9026**	0.5237	0.1482
ICLTD-BN	**0.6028**	**0.9214**	**0.9151**	0.6077	0.7296	0.3461
ICLTD-w/o-LSSC	0.1187	0.4267	0.7361	0.4559	0.5215	0.2742
ICLTD	0.4417	0.4125	0.8994	0.7838	**0.7698**	**0.5748**

Table 8. AUC_{SNPR} results for the ablation studies using the MT-ABU datasets. The best results are indicated in bold.

Exp \ Dataset	A1	A2	B1	B2	U1	U2
ICLTD-w/o-ICLM	1.000	1.005	1.000	1.001	1.002	1.000
ICLTD-ZG	2.524	7.482	12.554	15.076	3.042	1.724
ICLTD-BN	40.127	29.447	47.522	38.923	23.485	30.188
ICLTD-w/o-LSSC	340.796	**1000.888**	**999.948**	1605.910	326.152	421.725
ICLTD	**408.216**	753.550	710.780	**1749.182**	**342.586**	**465.132**

After zeroing the gradient signal transmission in the ICLM, the performance of target–background separation is better than that of ICLTD-w/o-ICLM, which is reflected by the visualization and AUC_{TDBS} results. However, the $AUC_{(D,F)}$ and AUC_{SNPR} results of ICLTD-ZG are unsatisfactory, which means that ICLTD-ZG mistakenly detects many background samples. The comparison of ICLTD-w/o-ICLTD and ICLTD-ZG validates that changing the feature distribution with the ICLM could improve the target–background separation. The comparison of ICLTD-ZG and ICLTD-w/o-LSSC proves that the gradient transmission by the ICLM is the key to success.

The $AUC_{(D,F)}$ results after replacing the ICLM with the original BN are promising for all datasets. The target–background separation performance in terms of amplitude difference was excellent. However, the AUC_{SNPR} results and background suppression performance of the ICLTD-BN are not as good as those of the ICLTD-w/o-LSSC. Especially for the MT-ABU-A2, MT-ABU-B1, and MT-ABU-U1 datasets, the ICLTD-BN mistakenly detected a few background spectra. These results validate that insufficient prior spectral features for normalization reduce the regularization strength, resulting in unpromising background suppression performance.

After removing the LSSC, the completeness of targets in the visualization results decreased. The AUC_{TDBS} results of the ICLTD-w/o-LSSC also degraded. The performance comparison between the ICLTD-w/o-LSSC and ICLTD proves the effectiveness of the LSSC in improving target detectability. It is worth noting that the background suppression ability of the ICLTD does not decrease significantly with the addition of the LSSC. Therefore, the LSSC can utilize local spectral similarities to improve target–background separation performance.

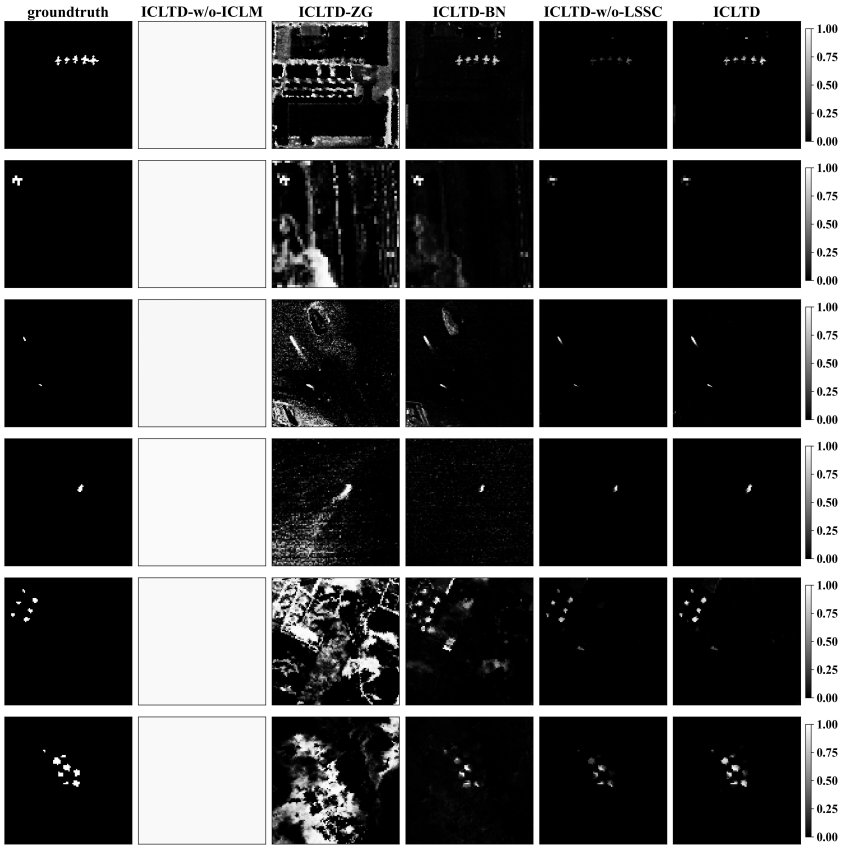

Figure 12. Ablation study results in terms of detection maps visualization. Due to model collapse, ICLTD-w/o-ICLM considered all the in-scene spectra as targets, resulting in poor contrast of its visualization results.

4. Conclusions

This paper introduced a DL-based method for HTD that effectively mitigates the challenge of spectral variability in multi-temporal HSIs. The proposed approach utilizes only prior spectra for supervised learning and does not rely on pseudo-data generation for optimization. By introducing implicit contrastive learning between the prior and in-scene spectra, the optimization process is regularized through a specifically designed ICLM. A theoretical analysis demonstrated the efficacy of the ICLM, showing that it successfully transfers differentiated gradient signals from prior spectral features to the representation of in-scene spectra, leveraging their inherent differences. This allows the ICLTD to learn discriminative representations of the in-scene spectra, avoiding over-fitting the prior spectra. Consequently, it is capable of effectively distinguishing variational in-scene target spectra from background spectra and adapting to the spectral variability in multi-temporal HSIs. Additionally, the proposed LSSC leverages the spectral similarity of multi-pixel targets within local neighborhoods to enhance target detectability. The performance of the proposed method was evaluated using three multi-temporal image sets obtained from AVIRIS data, which demonstrated its robustness under spectral variability. A comparison with classical detectors and DL detectors confirmed the superior performance

of the proposed method in achieving a balance between target detectability and background suppression. Based on the experimental results, practical recommendations for application of the ICLTD emphasize the consideration of adjusting the strength of implicit contrast learning based on the intensity of spectral variations. In future research, exploring how to better combine implicit contrastive learning with annotation-based supervised learning from both theoretical and experimental perspectives can enhance the performance of hyperspectral target detection and expand its application scenarios.

Author Contributions: X.Z. (Xiaodian Zhang) and J.W. conceived and designed the study. X.Z. (Xiaodian Zhang) constructed the model, implemented the experiments, and drafted the manuscript. Z.H. and Z.Y. contributed to improving the manuscript, and P.W. collected the hyperspectral datasets. K.G., W.L. and X.Z. (Xiaobin Zhao) provided the overall guidance to this work and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Science and Technology Ministry of China under Grant 2022YFC3301602, and in part by the National Natural Science Foundation of China under Grant U2241275.

Data Availability Statement: Code and data will be available at https://github.com/zxd52csx/Implicit-constrastive-learning-based-HTD, accessed on 23 December 2023.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Makki, I.; Younes, R.; Francis, C.; Bianchi, T.; Zucchetti, M. A survey of landmine detection using hyperspectral imaging. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 40–53. [CrossRef]
2. Adep, R.N.; Ramesh, H. EXhype: A tool for mineral classification using hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 106–118. [CrossRef]
3. Lin, C.; Chen, S.Y.; Chen, C.C.; Tai, C.H. Detecting newly grown tree leaves from unmanned-aerial-vehicle images using hyperspectral target detection techniques. *ISPRS J. Photogramm. Remote Sens.* **2018**, *142*, 174–189. [CrossRef]
4. Borsoi, R.A.; Imbiriba, T.; Bermudez, J.C.M.; Richard, C.; Chanussot, J.; Drumetz, L.; Tourneret, J.Y.; Zare, A.; Jutten, C. Spectral Variability in Hyperspectral Data Unmixing: A comprehensive review. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 223–270. [CrossRef]
5. Axelsson, M.; Friman, O.; Haavardsholm, T.V.; Renhorn, I. Target detection in hyperspectral imagery using forward modeling and in-scene information. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 124–134. [CrossRef]
6. Kruse, F.A.; Lefkoff, A.; Boardman, J.; Heidebrecht, K.; Shapiro, A.; Barloon, P.; Goetz, A. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [CrossRef]
7. Chang, C.I. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Trans. Inf. Theory* **2000**, *46*, 1927–1932. [CrossRef]
8. Kelly, E.J. An adaptive detection algorithm. *IEEE Trans. Aerosp. Electron. Syst.* **1986**, *AES-22*, 115–127. [CrossRef]
9. Kraut, S.; Scharf, L.L. The CFAR adaptive subspace detector is a scale-invariant GLRT. *IEEE Trans. Signal Process.* **1999**, *47*, 2538–2541. [CrossRef]
10. Kraut, S.; Scharf, L.L.; Butler, R.W. The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic. *IEEE Trans. Signal Process.* **2005**, *53*, 427–438. [CrossRef]
11. Farrand, W.H.; Harsanyi, J.C. Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique. *Remote Sens. Environ.* **1997**, *59*, 64–76. [CrossRef]
12. Chang, C.I. Hyperspectral Target Detection: Hypothesis Testing, Signal-to-Noise Ratio, and Spectral Angle Theories. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–23. [CrossRef]
13. Jiao, X.; Chang, C.I. Kernel-based constrained energy minimization (K-CEM). In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV. SPIE, Orlando, FL, USA, 16–20 March 2008 ; Volume 6966, pp. 523–533.
14. Zou, Z.; Shi, Z. Hierarchical suppression method for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 330–342. [CrossRef]
15. Zhao, R.; Shi, Z.; Zou, Z.; Zhang, Z. Ensemble-based cascaded constrained energy minimization for hyperspectral target detection. *Remote Sens.* **2019**, *11*, 1310. [CrossRef]
16. Chen, Z.; Lu, Z.; Gao, H.; Zhang, Y.; Zhao, J.; Hong, D.; Zhang, B. Global to local: A hierarchical detection algorithm for hyperspectral image target detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5544915. [CrossRef]
17. Zhao, X.; Hou, Z.; Wu, X.; Li, W.; Ma, P.; Tao, R. Hyperspectral target detection based on transform domain adaptive constrained energy minimization. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102461. [CrossRef]
18. Zhu, D.; Du, B.; Zhang, L. Learning Single Spectral Abundance for Hyperspectral Subpixel Target Detection. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1 –11. [CrossRef] [PubMed]

19. Zhu, D.; Du, B.; Hu, M.; Dong, Y.; Zhang, L. Collaborative-guided spectral abundance learning with bilinear mixing model for hyperspectral subpixel target detection. *Neural Netw.* **2023**, *163*, 205–218. [CrossRef] [PubMed]
20. Zhang, Y.; Du, B.; Zhang, L. A sparse representation-based binary hypothesis model for target detection in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 1346–1354. [CrossRef]
21. Li, W.; Du, Q.; Zhang, B. Combined sparse and collaborative representation for hyperspectral target detection. *Pattern Recognit.* **2015**, *48*, 3904–3916. [CrossRef]
22. Guo, T.; Luo, F.; Zhang, L.; Tan, X.; Liu, J.; Zhou, X. Target detection in hyperspectral imagery via sparse and dense hybrid representation. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 716–720. [CrossRef]
23. Zhao, X.; Li, W.; Zhao, C.; Tao, R. Hyperspectral Target Detection Based on Weighted Cauchy Distance Graph and Local Adaptive Collaborative Representation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5527313. [CrossRef]
24. Zhao, X.; Liu, K.; Gao, K.; Li, W. Hyperspectral time-series target detection based on spectral perception and spatial-temporal tensor decomposition. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5520812. [CrossRef]
25. Bhatti, U.A.; Yu, Z.; Chanussot, J.; Zeeshan, Z.; Yuan, L.; Luo, W.; Nawaz, S.A.; Bhatti, M.A.; Ain, Q.U.; Mehmood, A. Local similarity-based spatial–spectral fusion hyperspectral image classification with deep CNN and Gabor filtering. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5514215. [CrossRef]
26. Meng, S.; Wang, X.; Hu, X.; Luo, C.; Zhong, Y. Deep learning-based crop mapping in the cloudy season using one-shot hyperspectral satellite imagery. *Comput. Electron. Agric.* **2021**, *186*, 106188. [CrossRef]
27. Dou, Z.; Gao, K.; Zhang, X.; Wang, H.; Wang, J. Hyperspectral unmixing using orthogonal sparse prior-based autoencoder with hyper-Laplacian loss and data-driven outlier detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6550–6564. [CrossRef]
28. Li, W.; Wu, G.; Du, Q. Transferred deep learning for hyperspectral target detection. In Proceedings of the 2017 IEEE International Geoscience and Remote Sens. Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 5177–5180.
29. Zhu, D.; Du, B.; Zhang, L. Two-stream convolutional networks for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6907–6921. [CrossRef]
30. Gao, Y.; Feng, Y.; Yu, X.; Mei, S. Robust Signature-Based Hyperspectral Target Detection Using Dual Networks. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5500605 . [CrossRef]
31. Rao, W.; Gao, L.; Qu, Y.; Sun, X.; Zhang, B.; Chanussot, J. Siamese Transformer Network for Hyperspectral Image Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5526419. [CrossRef]
32. Zhang, X.; Gao, K.; Wang, J.; Hu, Z.; Wang, H.; Wang, P. Siamese Network Ensembles for Hyperspectral Target Detection with Pseudo Data Generation. *Remote Sens.* **2022**, *14*, 1260. [CrossRef]
33. Wang, Y.; Chen, X.; Wang, F.; Song, M.; Yu, C. Meta-Learning based Hyperspectral Target Detection using Siamese Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5527913. [CrossRef]
34. Jiao, J.; Gong, Z.; Zhong, P. Triplet Spectral-Wise Transformer Network for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5519817. [CrossRef]
35. Zhang, G.; Zhao, S.; Li, W.; Du, Q.; Ran, Q.; Tao, R. HTD-net: A deep convolutional neural network for target detection in hyperspectral imagery. *Remote Sens.* **2020**, *12*, 1489. [CrossRef]
36. Rao, W.; Qu, Y.; Gao, L.; Sun, X.; Wu, Y.; Zhang, B. Transferable network with Siamese architecture for anomaly detection in hyperspectral images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102669. [CrossRef]
37. Gao, Y.; Feng, Y.; Yu, X. Hyperspectral Target Detection with an Auxiliary Generative Adversarial Network. *Remote Sens.* **2021**, *13*, 4454. [CrossRef]
38. Wang, Y.; Chen, X.; Zhao, E.; Song, M. Self-supervised Spectral-level Contrastive Learning for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5510515. [CrossRef]
39. Zhang, X.; Gao, K.; Wang, J.; Hu, Z.; Wang, H.; Wang, P.; Zhao, X.; Li, W. Self-supervised learning with deep clustering for target detection in hyperspectral images with insufficient spectral variation prior. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *122*, 103405. [CrossRef]
40. Taghipour, A.; Ghassemian, H. Unsupervised hyperspectral target detection using spectral residual of deep autoencoder networks. In Proceedings of the 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA), Tehran, Iran, 6–7 March 2019; pp. 52–57.
41. Xie, W.; Zhang, X.; Li, Y.; Wang, K.; Du, Q. Background learning based on target suppression constraint for hyperspectral target detection. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2020**, *13*, 5887–5897. [CrossRef]
42. Gao, L.; Wang, D.; Zhuang, L.; Sun, X.; Huang, M.; Plaza, A. BS 3 LNet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5504218.
43. Li, Y.; Shi, Y.; Wang, K.; Xi, B.; Li, J.; Gamba, P. Target detection with unconstrained linear mixture model and hierarchical denoising autoencoder in hyperspectral imagery. *IEEE Trans. Image Process.* **2022**, *31*, 1418–1432. [CrossRef]
44. Xie, W.; Lei, J.; Yang, J.; Li, Y.; Du, Q.; Li, Z. Deep latent spectral representation learning-based hyperspectral band selection for target detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2015–2026. [CrossRef]
45. Shi, Y.; Li, J.; Li, Y.; Du, Q. Sensor-independent hyperspectral target detection with semisupervised domain adaptive few-shot learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6894–6906. [CrossRef]
46. Shi, Y.; Lei, J.; Yin, Y.; Cao, K.; Li, Y.; Chang, C.I. Discriminative feature learning with distance constrained stacked sparse autoencoder for hyperspectral target detection. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1462–1466. [CrossRef]

47. Shi, Y.; Li, J.; Yin, Y.; Xi, B.; Li, Y. Hyperspectral target detection with macro-micro feature extracted by 3-D residual autoencoder. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2019**, *12*, 4907–4919. [CrossRef]
48. Qin, H.; Xie, W.; Li, Y.; Jiang, K.; Lei, J.; Du, Q. Weakly supervised adversarial learning via latent space for hyperspectral target detection. *Pattern Recognit.* **2023**, *135*, 109125. [CrossRef]
49. Xie, W.; Yang, J.; Lei, J.; Li, Y.; Du, Q.; He, G. SRUN: Spectral regularized unsupervised networks for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 1463–1474. [CrossRef]
50. Xie, W.; Zhang, J.; Lei, J.; Li, Y.; Jia, X. Self-spectral learning with GAN based spectral–spatial target detection for hyperspectral image. *Neural Netw.* **2021**, *142*, 375–387. [CrossRef] [PubMed]
51. Abe Fetterman, J.A. Understanding Self-Supervised and Contrastive Learning with "Bootstrap Your Own Latent" (BYOL). 2020. Available online: https://imbue.com/research/2020-08-24-understanding-self-supervised-contrastive-learning/ (accessed on 23 December 2023).
52. Manolakis, D.; Marden, D.; Shaw, G.A. Hyperspectral image processing for automatic target detection applications. *Linc. Lab. J.* **2003**, *14*, 79–116.
53. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning. pmlr, Lille, France, 7–9 July 2015; pp. 448–456.
54. Chang, C.I. An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5131–5153. [CrossRef]
55. Hu, Z.; Gao, K.; Zhang, X.; Wang, J.; Wang, H.; Yang, Z.; Li, C.; Li, W. EMO2-DETR: Efficient-Matching Oriented Object Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5616814. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.