Learning a Cross-Modality Anomaly Detector for Remote Sensing Imagery

Jingtao Li, Xinyu Wang, *Member, IEEE*, Hengwei Zhao, *Member, IEEE*, and Yanfei Zhong, *Senior Member, IEEE*

***Abstract*— Remote sensing anomaly detector can find the objects deviating from the background as potential targets for Earth monitoring. Given the diversity in earth anomaly types, designing a transferring model with cross-modality detection ability should be cost-effective and flexible to new earth observation sources and anomaly types. However, the current anomaly detectors aim to learn the certain background distribution, the trained model cannot be transferred to unseen images. Inspired by the fact that the deviation metric for score ranking is consistent and independent from the image distribution, this study exploits the learning target conversion from the varying background distribution to the consistent deviation metric. We theoretically prove that the large-margin condition in labeled samples ensures the transferring ability of learned deviation metric. To satisfy this condition, two large margin losses for pixel-level and feature-level deviation ranking are proposed respectively. Since the real anomalies are difficult to acquire, anomaly simulation strategies are designed to compute the model loss. With the large-margin learning for deviation metric, the trained model achieves cross-modality detection ability in five modalities—hyperspectral, visible light, synthetic aperture radar (SAR), infrared and low-light—in zero-shot manner.**

***Index Terms*— Anomaly detection, remote sensing, transferability, cross-modality, cross-scene, unified detector.**

I. Introduction

REMOTE sensing images can be used to monitor anomalies on the Earth's surface in a large-scale and consistent space [1]. Anomaly detection in remote sensing (ADRS) task aims to find the pixels deviating from the background spectrally or spatially, which are detected without any prior knowledge [2], [3], [4]. The anomalies vary in category and electromagnetic response. For example, landslide anomalies exhibit a response in the visible and radar range, while fire anomalies are mainly related to the thermal infrared spectra [5]. Given the diversity in anomaly types and responses across modalities, building a transferring model with cross-modality detection ability for ADRS (as Fig. 1) to different modalities would be cost-effective and allow easy adaptation to new data sources and anomaly types.

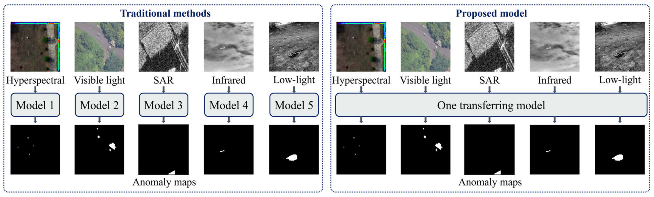


Fig. 1. The cross-modality detection paradigm of proposed model. Given the modalities with different imaging mechanisms and huge distribution difference, traditional models need to be trained for each modality while proposed model can infer the unseen modalities directly with zero-shot transferring ability.

Designing a such transferring model is challenging due to the difference in imaging mechanisms of different modalities. Specifically, the hyperspectral modality can record a continuous spectrum from visible to short-wave infrared [6], and thus the acquired imagery always has hundreds of channels for precise recognition [7], [8]. In contrast, the synthetic aperture radar (SAR) modality is a side-looking radar that records the received echoes coherently [9], [10], providing more structural information with several channels. Besides, large-scale scenes encompass diverse backgrounds, including forests, urban areas, and oceans, with highly variable distributions [11], [12].

Since the huge distribution difference, most anomaly detectors are still limited to a single modality since they aim to learn the certain background distribution for each image. They focus on describing the background distribution with a statistical-based [13], [14], [15], [16], representation-based [17], [18], [19], [20], or deep learning based method [21], [22], [23], [24] first, and then use some deviation metric directly to obtain the anomaly score. The statistics-based methods describe the background distribution with some statistical model (e.g., multivariate Gaussian distribution) [13]. The representation-based methods describe the background with a hand-crafted dictionary considering the low-rank and sparsity priors [2], [25]. The deep learning based models mostly use reconstruction models to learn the background distribution and assume that the normal pixels have a smaller reconstruction error than the anomaly ones [26], [27], [28]. After obtaining the background distribution, some deviation metric such as the Mahalanobis distance [15] and the mean

Received 27 February 2024; revised 5 September 2024 and 24 October 2024; accepted 25 October 2024. Date of publication 8 November 2024; date of current version 25 November 2024. This work was supported by the National Natural Science Foundation of China under Grant No.42325105. The associate editor coordinating the review of this article and approving it for publication was Prof. Gabriele Cavallaro. *(Corresponding author: Xinyu Wang.)*

Jingtao Li, Hengwei Zhao, and Yanfei Zhong are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing and Hubei Provincial Engineering Research Center of Natural Resources Remote Sensing Monitoring, Wuhan University, Wuhan 430072, China (e-mail: jingtaoli@whu.edu.cn; whu_zhaohw@whu.edu.cn; zhongyanfei@whu.edu.cn).

Xinyu Wang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China (e-mail: wangxinyu@whu.edu.cn).

Digital Object Identifier 10.1109/TIP.2024.3490894

squared error [29] is used directly to obtain the anomaly score. However, the background distribution always varies in unseen images and thus the prior constructed detector for certain background is not applicable any more. This is the main reason why the existing models need to be constructed again for each image and lack the cross-modality transferring ability.

To solve the transferring problem, finding an image-independent learning target is the core step. We observe that although the modality and scene have changed, most detectors use fixed deviation metrics (e.g., Mahalanobis distance) [15], [30] to compute the anomaly score. The learned background distribution act as the varying input for the deviation metric while the deviation metric itself is unchanged and image-independent. Inspired by this, we exploit to learn the deviation metric directly, which accepts the original image as input and ranks the deviation degree for each pixel. Different from the hand-crafted metrics, our score process is end-to-end without the need to obtain the background distribution first.

In this study, we build a cross-modality detector by learning an image-independent deviation metric. Instantiating the deviation metric as a learned deep model, we first theoretically prove that although the cross-modality images may be unseen at training stage, once the trained model can meet the large margin condition in the limited training samples, it can also rank the unseen anomaly and background correctly. Based on the proved Theorem, two large-margin deviation ranking losses are further proposed at pixel-level and feature-level. The pixel-level loss is derived from the common ranking metric (Area Under the Curve) AUC, and thus has a smaller gap between the optimization and the evaluation. The feature-level loss is designed to optimize the ranking of features with the hypersphere centers. Both the pixel-level and feature-level losses punish the small margin even for the correct ranking. Since the real anomalies are difficult to acquire, the anomaly simulation strategy is proposed to generate labeled anomalies and compute the large-margin losses.

In brief, the main contributions of this paper can be summarized as follows.

(1) The anomaly detection model with cross-modality transferring ability is built by converting the learning target from the certain background distribution to the image-independent deviation metric.
(2) We theoretically prove that meeting the large margin condition in training samples can guarantee the correct deviation rank for unseen anomaly and background.
(3) The large-margin ranking losses at pixel-level and feature-level are designed. The losses work together with simulated samples and punish the small margin even for the correct ranking.

The rest of this paper is organized as follows. Section II introduces the related work in remote sensing anomaly detection. Section III provides a detailed description of the motivation and the learning method of deviation metric. Section IV gives the experimental results and analysis. Finally, the paper is concluded in Section V. The data and code are available at http://rsidea.whu.edu.cn/deviation_AD.htm.

II. Related Work

A. Anomaly Detection Task in Remote Sensing

ADRS involves finding the objects that are anomalous to the background, without any prior information [31]. There is not an unambiguous way to define an anomaly, which is generally identified as an observation deviating from the background, spectrally or spatially [2], [4]. In fact, the category of the anomalies depends on the particular application. The anomalies can be the camouflage [43] or vehicles in military surveillance [32], rare minerals in geological detection [33], infected trees in forestry [30], and ships on the sea [33]. Since the ADRS methods do not use any prior knowledge, they cannot distinguish between real anomalies and detections that are not of interest. The detection result is often a first step, which provides the potential targets for the subsequent recognition [34].

Some fields may seem similar to the ADRS methods, but there are significant differences. Anomaly detection in medical or industrial images finds the anomaly pattern given a set of normal samples [35], where the normal pattern is no longer the background defined in the ADRS. The detected anomalies have both large and small areas. Despite some researchers having defined the normal pattern as the same as the industrial one in high-resolution optical images [30], we inherit te classical anomaly definition in the remote sensing community and treat the background as the normal pattern in each scene. Compared to tiny object detection [36], [37], the ADRS task is unsupervised without preset categories and labeled training samples. In addition, the anomalies in an ADRS are always small and rare, while tiny object detection also considers abundant small objects (e.g., cars in a parking lot).

B. Anomaly Detection Methods in Remote Sensing

Since the difficulty to acquire the real anomalies, most ADRS methods aim to extract the discriminative background features first and then use a distance metric to assign the anomaly score for each pixel. According to the principles of background learning, the detection models can be divided into three categories: statistical-based [13], [14], [15], [16], [38], [39] models, representation-based [17], [18], [19], [20] models, and deep learning based method [21], [22], [23], [24], [40], [41].

1) Statistics-Based Models: This statistical models aim to describe the background distribution with statistical techniques [14], where the likelihood implies the anomaly degree. For example, the classic Reed-Xiaoli (RX) detector models the background as a multivariate Gaussian distribution [15]. The Mahalanobis distance between the test pixel and the modeled distribution is then treated as the anomaly degree. Inspired by the RX detector, many improved variants have been proposed, such as the kernel RX-AD [16], weighted-RX-AD and linear filter-based RX-AD [42] and spectral-spatial feature extraction-based AD [43]. Recently, Chang proposed [44], [45] a target-to-anomaly conversion mechanism, which converts many well-known target detectors to the corresponding anomaly versions. Except for the accuracy improvement, some researchers have focused on real-time processing with RX detectors [38], [46], [47]. To address the difficulty of determining the distribution form, statistical cluster centers and decision hyperspheres have also been deployed [48], [49]. For the SAR modality, Haitman et al. [50] used both the RX detector and the non-negative matrix factorization (NNMF) learning algorithm to detect the sludge pools in Israel. Despite the statistical methods having a clear mathematical basis, the constructed distribution is only suitable for single images [27] and does not have the ability to be cross-modal or cross-scene.

2) Representation-Based Models: The representation-based models construct the detector considering the prior properties of the anomalies or background [17], [51], and include sparsity, collaborative, and low-rank based detectors. Ling et al. [18] constructed a sparsity-based detector with the sum-to-one and non-negativity constraints, making the detector less sensitive to the anomalies. Differing from sparse representation, collaborative representation assumes that the background pixels can be reconstructed by the surrounding pixels while the anomalies cannot [19]. The classic collaborative representation detector (CRD) follows this assumption [20]. To make full use of the global structural information (i.e., low-rank property), the low-rank and sparse matrix decomposition model (LSDM) was designed by decomposing the hyperspectral image into a low-rank background and sparse anomalies [17]. Sun et al. [52] implemented the LSDM technique with robust principal component analysis (RPCA) [53]. Zhang et al. [54] proposed a detector based on the low-rank and sparse matrix decomposition (LRaSMD) technique and applied the Mahalanobis distance to estimate the background part (LSMAD). Xu et al. [55] first introduced the background dictionary and proposed a detector based on low-rank and sparse representation (LRASR). Although the representation models do not rely on specific statistical distribution, the used background dictionary needs to be constructed for each modality and scene, limiting the transferring ability.

3) Deep Learning Based Models: Most deep learning based models follow a two-step paradigm [27], [56], where they assume that the normal pixels have a smaller reconstruction error with the deep model than the anomaly ones. Li et al. [40] first introduced a convolutional neural network (CNN) into the hyperspectral anomaly detection (HAD) task in a supervised way. To detect anomalies according to a practical situation, some unsupervised methods have been proposed. For example, Xie et al. [23] proposed the spectral constrained adversarial autoencoder (SC_AAE), where a spectral constraint strategy is incorporated for better latent representation. However, these methods always involve complicated manual parameter setting and preprocessing steps. To this end, Wang et al. [29] proposed the autonomous hyperspectral anomaly detection network (Auto-AD) with an adaptive-weighted loss function, where a high reconstruction error implies anomaly. Except for the autoencoder model, generative adversarial network (GAN)-based models have also been used, where the generation error from real images is treated as the anomaly degree [57], [58], [59]. For example, Jiang et al. [59] introduced a semi-supervised GAN with dual RX detector to learn the discriminative reconstruction of background and anomalies. Inspired by the fact that both the autoencoder-based models and GAN-based models adopt the reconstruction proxy task and need to be trained for each image, Li et al. [27] proposed the one-step detection paradigm and transferred direct detection (TDD) model, where the proxy task is abandoned and the trained model can be transferred to unseen images directly. However, the TDD model is still limited in the hyperspectral modality due to the proxy classification optimization and the simulated spectral anomalies.

4) Fast Anomaly Detection: Since anomalies may appear in a short time and bring huge losses, many efforts have been made to improve the detection speed. Chen et al. [38] designed causal processing with Kalman filters and achieved real-time performance. To better conform to the push-broom scanners, Díaz et al. [60] proposed a line-by-line anomaly detector (LbL-FAD), which used hardware-friendly alternative to compute the orthogonal subspace spanned by selected background pixels to make the anomalies easily separated. López-Fandiño et al. [61] designed a parallel algorithm to be executed on multi-node heterogeneous computing platforms based on Reed–Xiaoli (RX) [15]. A et al. [62] proposed a fast local RX (FLRX) detector to achieve near-time performance. It can be seen that most fast detectors are based on statistical models to achieve real-time performance. Although prior work has used field programmable gate array (FPGA) to speed up the deep anomaly detectors for multispectral imagery [63], their transferability is still limited in known scenes due to the learning target of certain background. In hyperspectral community, the paradigm of training and testing on each image prevents the deep model from being real-time. This study tackles the problem by transforming the learning target from varying background to fixed deviation ranking, eliminating the training time for fast processing on unseen images.

III. A Transferring Model for Remote Sensing Anomaly Detection

In this section, we first formulate the ADRS task and clarify the motivation in Section III-A, where our main idea is to change the learning target from the varying image distribution to the image-independent deviation metric (Fig. 2). The analysis in Section III-B shows that satisfying the large margin condition in the labeled samples is the key for the transferring ability of learned deviation metric. To satisfy the condition, two large-margin losses are proposed in pixel-level and feature-level respectively for the correct deviation ranking in Section III-C. Since real anomalies are difficult to acquire, we design the anomaly simulating strategies in Section III-D for computing the deviation ranking loss. Fig. 3 gives an overview of the built transferring model.

A. Motivation: From Single to Cross-Modality Detection

Given a remote sensing image $\mathbf{X} \in R^{H\times W\times C}$, $\mathbf{X} = \mathbf{B} + \mathbf{A}$ in the ideal condition without noise, where $\mathbf{B}$ is the background and $\mathbf{A}$ is the anomaly component. In ADRS task, $\mathbf{A}$ is always the small target with the empirical ratio in range [0.0019%, 0.48%] obtained by statistics of 12 well-known datasets from [29], [55], [64], [65], [66], [67], [68], [69], [70].

Regardless of the instantiation difference, current detection models rely on both the image distribution $P(\mathbf{X})$ and the deviation metric function S, which measures the scalar

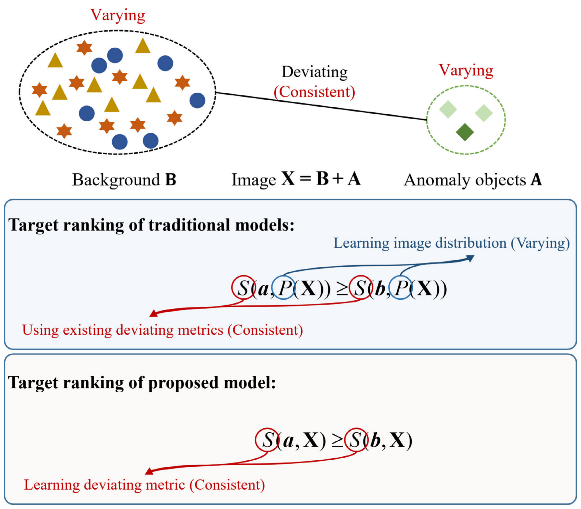

Varying
Deviating
(Consistent)
Varying
Background **B**
Image **X = B + A**
Anomaly objects **A**
Target ranking of traditional models:
Learning image distribution (Varying)
$S(\boldsymbol{a}, P(\mathbf{X})) \geq S(\boldsymbol{b}, P(\mathbf{X}))$
Using existing deviating metrics (Consistent)
Target ranking of proposed model:
$S(\boldsymbol{a}, \mathbf{X}) \geq S(\boldsymbol{b}, \mathbf{X})$
Learning deviating metric (Consistent)

Fig. 2. Description of our main principle difference compared to traditional models. Traditional models focus on learning the certain image distribution first and then use some existing deviating metric to rank the anomaly score. In different modalities, since the distributions of background and anomaly are varying, the prior learned model cannot be transferred to unseen image distribution. Inspired by the fact that the deviating metric S is independent and consistent for all the modality image, proposed model aims to bypass the image distribution learning and learn the deviating metric directly, achieving the cross-modality detection.

deviation degree of the pixel $\mathbf{x} \in \mathbf{X}$ and the $P(\mathbf{X})$. The deviation degree for each pixel reflects its occurrence probability and the distribution difference with the whole image. Ideally, any anomaly pixel $\boldsymbol{a} \in \mathbf{A}$ and any background pixel $\boldsymbol{b} \in \mathbf{B}$ should satisfy the ranking inequality $S(\boldsymbol{a}, P(\mathbf{X})) \geq S(\boldsymbol{b}, P(\mathbf{X}))$, implying the higher deviation and anomaly score of $\mathbf{A}$ than $\mathbf{B}$. For example, RXD instantiates the $P(\mathbf{X})$ as the multivariate Gaussian distribution and instantiates the S as the Mahalanobis distance [15]. LRASR [55] instantiates the $P(\mathbf{X})$ as a background dictionary and instantiates the S as the reconstruction error.

We observe that most models only concern about the quality of $P(\mathbf{X})$ and use some existing distance metric directly to get the final anomaly map (e.g., Mahalanobis distance). They mainly differ in the methods of representing $P(\mathbf{X})$ (statistical-based, representation-based or deep learning-based). However, when $P(\mathbf{X})$ has changed given an unseen image of different modality, the model needs to be rebuilt or trained, lacking the transferring detection ability.

To solve the transferring problem and increase the flexibility of the detection model, ***our main idea is to abandon the learning target of*** $P(\mathbf{X})$ ***but learn the deviation metric*** S ***directly***. Since S is independent of the $\mathbf{X}$, the learned S can be consistent given any unseen image, thus achieving the cross-modality transferring detection.

B. Learning Deviation Metric for Correct Ranking

We made S learnable by instantiating it as a trained deep model. The expected S can score the deviation degree for each pixel and satisfy the ranking inequality $S(\boldsymbol{a}, \mathbf{X}) \geq S(\boldsymbol{b}, \mathbf{X})$. Different from the traditional deviation metric such as the Mahalanobis distance, our S does not need to obtain the $P(\mathbf{X})$ first. Fig. 2 shows our main difference compared to the traditional models. For simplicity, unless otherwise specified, $S(\boldsymbol{a}, \mathbf{X})$ is shortened to $S(\boldsymbol{a})$ and $S(\boldsymbol{b}, \mathbf{X})$ is shortened to $S(\boldsymbol{b})$ in the following paragraphs.

To answer the question of " ***how to ensure the learned deviation ranking ability of* S *transferring***?". In this section, we theoretically prove that the transferring ability of S can be achieved once meeting the large margin condition in limited labeled samples (statement is provided in Theorem 1).

Theorem 1: Set Q_l be the training set of many labeled samples (the anomaly pixel $\boldsymbol{a}_j \in R^C$ indexed by j, the background pixel $\boldsymbol{b}_i \in R^C$ indexed by i). Set δ be the smallest radius, such that for any unseen anomaly pixel $\boldsymbol{u}_a$ or the unseen background pixel $\boldsymbol{u}_b$, $\boldsymbol{u}_a$ is in the δ-ball of some $\boldsymbol{a}_j$ in Q_l and $\boldsymbol{u}_b$ is in the δ-ball of some $\boldsymbol{b}_i$ in Q_l. If the score function S meets the λ_s-Lipschitz continuous condition and has correctly ranked the Q_l with a large margin, i.e., $S(\boldsymbol{a}_j) - S(\boldsymbol{b}_i) \geq 2\delta\lambda_s$ holds for all the labeled pixels, then S can also rank the unseen pixels of different modality correctly, i.e., $S(\boldsymbol{u}_a) \geq S(\boldsymbol{u}_b)$.

Proof: Considering the λ_s-Lipschitz continuous property of S and the $\boldsymbol{u}_a$ and $\boldsymbol{u}_b$ are assumed to be close to $\boldsymbol{a}_j$ and $\boldsymbol{b}_i$ respectively with the distance smaller than δ. Eq. (1) and Eq. (2) can be obtained.

$$S(\boldsymbol{a}_j) - \delta\lambda_s \leq S(\boldsymbol{u}_a) \tag{1}$$

$$-S(\boldsymbol{b}_i) - \delta\lambda_s \leq -S(\boldsymbol{u}_b) \tag{2}$$

Adding the inequalities (1) and (2), and with the condition $S(\boldsymbol{a}_j) - S(\boldsymbol{b}_i) \geq 2\delta\lambda_s$, we can further obtain the Eq. (3). Thus, $S(\boldsymbol{u}_a) \geq S(\boldsymbol{u}_b)$ can hold.

$$0 \leq S(\boldsymbol{a}_j) - S(\boldsymbol{b}_i) - 2\delta\lambda_s \leq S(\boldsymbol{u}_a) - S(\boldsymbol{u}_b) \tag{3}$$

Theorem 1. shows that if the learned S satisfies the Lipschitz continuous condition and the large margin condition in labeled samples of Q_l, it can also rank the unseen anomalies and background correctly and thus achieve the transferring ability. Lipschitz continuous is a common condition and controlled by the regularization strength of the deep model. Thus, meeting the large margin condition in labeled samples is the key for guaranteeing the transferring ability.

For the deviation metric learning of ADRS, we meet the large margin condition in both pixel-level and feature-level optimization. The pixel-level loss is optimized directly for the deviation ranking metric (i.e., AUC), where the discrete zero-one loss is replaced with the designed differentiable log loss. Even a correct ranking has been obtained, the penalty exists and changes according to the margin. The feature-level loss optimizes the deviation ranking of extracted features in an equivalent way, which enlarges the distance of the hypersphere centers between the anomaly and background features while decreasing their hypersphere radiuses at the same time. Both pixel-level and feature-level losses work together to strength the large margin ranking learning.

Besides, since ADRS task is unsupervised without real anomalies, we propose an anomaly generating strategy to generate the paired labeled samples by simulating the deviation ranking relationship. The simulated samples convert ADRS from the unsupervised learning setting into the pseudo supervised setting, which are used to compute the pixel-level and feature-level ranking margin losses.

Optimized with the large margin losses (Section III-C) and the simulated anomaly samples (detailed in Section III-D),

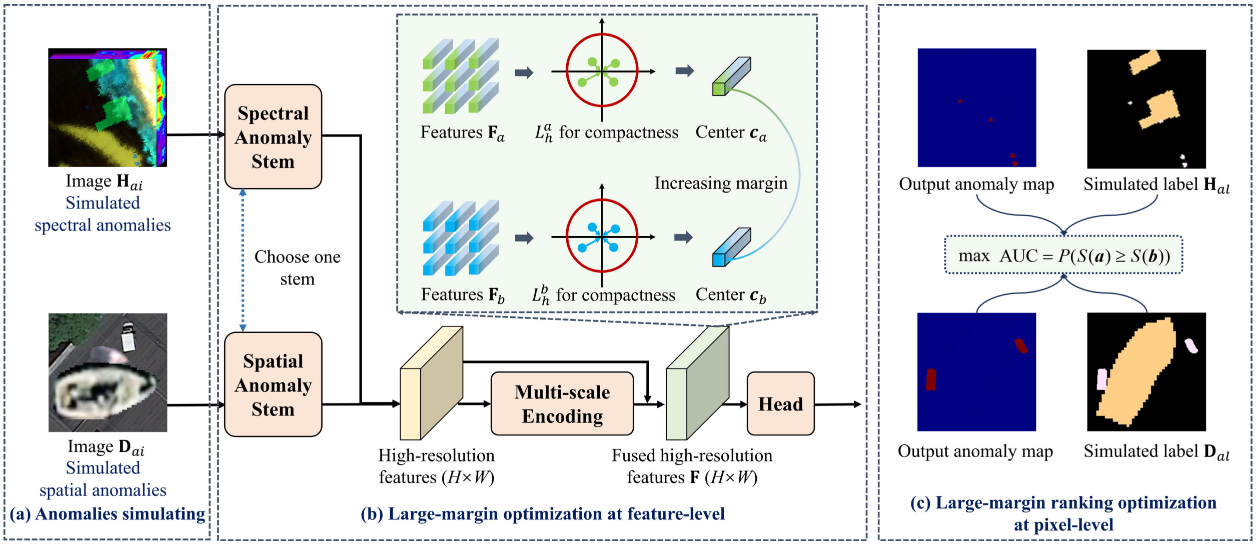


Image $\mathbf{H}_{ai}$
Simulated spectral anomalies
Image $\mathbf{D}_{ai}$
Simulated spatial anomalies
(a) Anomalies simulating
Spectral Anomaly Stem
Choose one stem
Spatial Anomaly Stem
Features $\mathbf{F}_a$
L_h^a for compactness
Center $\boldsymbol{c}_a$
Increasing margin
Features $\mathbf{F}_b$
L_h^b for compactness
Center $\boldsymbol{c}_b$
High-resolution features ($H\times W$)
Multi-scale Encoding
Fused high-resolution features $\mathbf{F}$ ($H\times W$)
Head
(b) Large-margin optimization at feature-level
Output anomaly map
Simulated label $\mathbf{H}_{al}$
max AUC = $P(S(\boldsymbol{a}) \geq S(\boldsymbol{b}))$
Output anomaly map
Simulated label $\mathbf{D}_{al}$
(c) Large-margin ranking optimization at pixel-level

Fig. 3. The built transferring detection model with learning the deviation metric for correct ranking. According to the proven Therorem1, meeting the large-margin condition in labeled samples is the key to ensure the transferring ability of the learned deviation metric. To learn the large margin ranking, we design the pixel-level and feature-level optimization, respectively. Optimization at pixel-level (c) optimize the ranking metric AUC directly, where the discrete zero-one loss is replaced by the designed surrogate loss to be differentiable and large margin (Section III-C). Optimization at feature-level (b) aims to enlarge the ranking margin of features, which decreases the hypersphere radiuses enclosing the anomaly and background features and also increases their center distance (Section III-C). Besides, since the real anomalies are difficult to acquire, we simulate both spectral and spatial anomalies (a) to compute the large-margin losses.

a transferring model for ADRS task can be built as in Fig. 3, which can be trained only once and transferred to unseen images of different modality directly.

C. Large-Margin Ranking Losses

Traditional ranking learning adopts proxy losses (e.g., cross entropy loss for the classification task [71]) or the discriminate ranking losses (e.g., the average precision (AP) loss) [72]. To be more consistent with the common ranking metric, we derive the pixel-level large-margin loss from the AUC directly and also design the feature-level loss to strengthen the large margin ranking learning.

1) Pixel-Level Ranking Loss: We derive the loss from the AUC metric to keep the optimization process and the ranking evaluation consistent. AUC measures the probability of that $\boldsymbol{a} \in \mathbf{A}$ will rank higher than $\boldsymbol{b} \in \mathbf{B}$ and can be written in the integral form as in Eqs. (4) and (5),

$$\mathrm{AUC} = P(S(\boldsymbol{a}) \geq S(\boldsymbol{b})) = \int_0^1 \mathrm{TPR@FPR}_\eta(S) d\eta \quad (4)$$

$$\mathrm{TPR@FPR}_\eta(S) = \max \mathrm{TPR}(S,t) \quad s.t.\ \mathrm{FPR}(S,t) \leq \eta \quad (5)$$

where TPR is the true positive rate and the FPR is the false positive rate. The anomaly is considered as the positive class while the background is negative. The changing false alarm rate η is decided by the corresponding threshold t, which transforms the continuous anomaly map into a binary map. The TPR and FPR can be expressed with the zero-one loss L_{01} (i.e., 0 for correct prediction and 1 for wrong prediction) as in Eqs. (6) and (7).

$$\mathrm{TPR}(S,t) = \frac{\sum_{\boldsymbol{a}\in\mathbf{A}} 1 - L_{01}(S(\boldsymbol{a}),t)}{|\mathbf{A}|} \quad (6)$$

$$\mathrm{FPR}(S,t) = \frac{\sum_{\boldsymbol{b}\in\mathbf{B}} 1 - L_{01}(S(\boldsymbol{b}),t)}{|\mathbf{B}|} \quad (7)$$

For the large-margin optimization, the sigmoid loss or the p-order hinge loss [73] can be used as the surrogate loss to make the discrete L_{01} differentiable. However, they always need the another hyperparameter to control the margin. To tackle this, we choose to achieve the margin optimization with the help of log curve rather than the hyperparameter. The proposed surrogate loss $\overline{L}(\boldsymbol{x},t)$ for L_{01} is defined in Eq. (8), which covers the four situations.

$$\overline{L}(\boldsymbol{x},t) = \begin{cases} -\log(S(\boldsymbol{x})) & if \boldsymbol{x} \in \mathbf{A} \text{ and } S(\boldsymbol{x}) \geq t \\ \dfrac{\log(S(\boldsymbol{x}))}{\log(t)} & if\ \boldsymbol{x} \in \mathbf{A} \text{ and } S(\boldsymbol{x}) < t \\ -\log(1-S(\boldsymbol{x})) & if\ \boldsymbol{x} \in \mathbf{B} \text{ and } S(\boldsymbol{x}) < t \\ \dfrac{\log(1-S(\boldsymbol{x}))}{\log(t)} & if\ \boldsymbol{x} \in \mathbf{B} \text{ and } S(\boldsymbol{x}) \geq t \end{cases} \quad (8)$$

For the first and third situations, although the model has already scored $\mathbf{A}$ and $\mathbf{B}$ correctly given the threshold t (i.e., $S(\boldsymbol{a}) \geq t$ or $S(\boldsymbol{b}) < t$), the loss exists and encourages the larger score margin. The smaller margin implies larger loss and the correlation is controlled by the log curve. If the model has given a wrong score ranking, the corresponding loss relies on both the score value and the degree of the threshold t. For example, when the t is very large near one, a lot of anomaly pixels would be classified wrongly as the background, resulting a large and unreasonable loss. To deal with this problem, we multiply the loss with a factor $1/\log(t)$, which gives less weight to unreasonable thresholds. Replacing the L_{01} in Eqs. (6) and (7) with $\overline{L}(\boldsymbol{x},t)$, we can get the surrogate ones denoted as $\overline{\mathrm{TPR}}(S,t)$ and $\overline{\mathrm{FPR}}(S,t)$ respectively as in Eq. (9).

$$\overline{\mathrm{TPR@FPR}_\eta}(S) = \max \overline{\mathrm{TPR}}(S,t)\ s.t.\ \overline{\mathrm{FPR}}(S,t) \leq \eta \quad (9)$$

TABLE I

THE DETAILED ARCHITECTURE AND FEATURE SHAPE OF SPATIAL AND SPECTRAL STEMS. THE CONVOLUTIONAL LAYER IS REPRESENTED AS CONV (INPUT CHANNEL, OUTPUT CHANNEL, KERNEL SIZE, STRIDE, PADDING) AND BN REPRESENTS THE BATCH NORMALIZATION

Input shape: H×W×C (interpolate images to same shape)		
Spatial stem architecture	Spectral stem architecture	Output shape
Conv(C,64,3,1,1)	Conv(C,64,1,1,0)	H×W×64
BN+Relu	BN+Relu	H×W×64
Conv(64,64,3,1,1)	Conv(64,64,1,1,0)	H×W×64
BN+Relu	BN+Relu	H×W×64

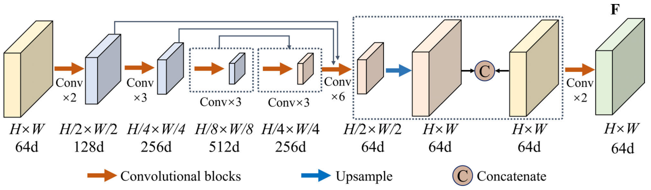


Fig. 4. The detailed architecture of multi-scale encoding process in Fig. 3(b). The usage feature color and size are consistent with Fig. 3(b). Conv-2 represents two cascaded convolution layers, and 64d represents that the channel number of corresponding feature cube is 64. The output fusing features **F** have the same spatial resolution ($H \times W$) with input image and is used to compute the high-resolution deviation score map in the convolutional head.

Theorem 2: The surrogate $\overline{\text{TPR@FPR}}_\eta(S)$ is a lower bound for the original $\text{TPR@FPR}_\eta(S)$.

Proof: Considering Eqs. (6) and (8), $L_{01}(S(\boldsymbol{x}) \geq t)$ is 0 but $0 \leq \overline{L}(\boldsymbol{x}, t) < 1$ when $\boldsymbol{x} \in \mathbf{A}$ and $S(\boldsymbol{x}) \geq t$. $L_{01}(S(\boldsymbol{x}) \geq t)$ is 1 but $\overline{L}(\boldsymbol{x}, t) > 1$ when $\boldsymbol{x} \in \mathbf{A}$ and $S(\boldsymbol{x}) < t$. Thus, $\overline{\text{TPR}}(S, t)$ with the $\overline{L}$ is the lower bound of the $\text{TPR}(S, t)$ with the L_{01}. Similarly, $\overline{\text{FPR}}(S, t)$ is the upper bound of the original $\text{FPR}(S, t)$ considering Eqs. (7) and (8) together. Therefore, $\overline{\text{TPR}}(S, t) \leq \text{TPR}(S, t)$ and $\overline{\text{FPR}}(S, t) \geq \text{FPR}(S, t)$, and the Theorem is proved.

Theorem 2 proves the surrogate rationality of the designed differentiable large-margin $\overline{L}(\boldsymbol{x}, t)$ for the discrete L_{01}. After replacing the $\text{TPR@FPR}_\eta(S)$ in Eq. (4) with $\overline{\text{TPR@FPR}}_\eta(S)$, we can use the Lagrange multiplier λ to deal with the constraint of $\overline{\text{FPR}}(S, t)$ and then approximate the integral in with a discrete sum over the anchor values.

The final obtained large-margin ranking loss at pixel-level L_p is given in Eq. (10), where k anchors exist in the range [0, 1], with each anchor corresponding to the false alarm rate η_i, threshold t_i, and multiplier Δ_i. $\Delta_i = \eta_i - \eta_{i-1}$ for $\forall\ i = 1 \ldots k$.

$$L_p = \min_{S, t_1, \ldots, t_k} \max_{\lambda_1, \ldots, \lambda_k} \sum_{i=1}^{k} \Delta_i (1 - \overline{\text{TPR}}(S, t_i)) + \lambda_i (\overline{\text{FPR}}(S, t_i) - \eta_i |A|) \tag{10}$$

2) Feature-Level Ranking Loss: Since the remote sensing anomalies are always tiny objects, the high-resolution features are essential for preventing the loss of details. As in Fig. 3(b), two separate stems are designed to process the spectral and spatial anomalies respectively first. Both stems consist of two cascaded convolution layers, where spectral stem uses kernel size 1×1 covering spectral dimension only and spatial stem uses 3×3 covering both spatial and spectral dimensions. All the images are interpolated to the same shape for the stem processing and Table I shows the detailed internal workflow (detailed setting of input shape is described in Section IV-A.3). The output high resolution features are then processed by multi-scale blocks (as in Fig. 4), where a maximum down-sampling rate of 8× is set to filter out small anomaly objects. Concatenating the output context features with the previous high-resolution ones from stems, fused $\mathbf{F} \in R^{H \times W \times L}$ can be obtained, which provides both pixel-level and context-level information for each object and helps computing the deviation score with convolutional head.

To strengthen the large-margin ranking, our feature-level optimization is conducted on the multi-scale fused features $\mathbf{F} \in R^{H \times W \times L}$ with the original image spatial size $H \times W$ and the feature dimension L. The anomaly features $\mathbf{F}_a$ and background features $\mathbf{F}_b$ are extracted from $\mathbf{F}$ according to the sample label. Specifically, we decrease the hypersphere radiuses enclosing the anomaly and background features and also increase their center margin.

Given $\mathbf{F}_a$, its hypersphere center $\boldsymbol{c}_a \in R^L$ can be computed as the mean value along the spatial dimension as in Eq. (11).

$$\boldsymbol{c}_a = \text{mean} \boldsymbol{f}_a, \boldsymbol{f}_a \in \mathbf{F}_a \tag{11}$$

To decrease the radius R_a in Eq. (12) while making the hypersphere including $\mathbf{F}_a$ as much as possible, the hypersphere optimization L_h^a for $\mathbf{F}_a$ is formulated as the minimization problem in Eq. (13). The optimization L_h^b for $\mathbf{F}_b$ can be obtained in the similar way.

$$R_a^2 = \underset{\boldsymbol{f}_a \in \mathbf{F}_a}{\text{mean}} (\|\boldsymbol{f}_a - \boldsymbol{c}_a\|^2) \tag{12}$$

$$L_h^a = R_a^2 + \beta \underset{f_a \in \mathbf{F}_a}{\text{mean}} (\max\{\|f_a - c_a\|^2 - R_a^2, 0\}) \tag{13}$$

L_h^a and L_h^b make the corresponding hyperspheres compact and the hypersphere centers can represent the overall feature distribution in a certain extent. With the constraints of L_h^a and L_h^b, the feature level loss L_f enlarges the ranking margin of $\mathbf{F}_a$ and $\mathbf{F}_b$ by increasing the distance of the anomaly hypersphere center $\boldsymbol{c}_a$ and the background hypersphere center $\boldsymbol{c}_b$. The L_f is formulated in Eq. (14). Since the three terms have the same order of magnitude and importance, the loss ratio is set 1:1:1.

$$L_f = -\|\boldsymbol{c}_b - \boldsymbol{c}_a\|^2 + L_h^a + L_h^b \tag{14}$$

In total, the pixel-level loss L_p and the feature-level loss L_f work together for the large-margin score ranking target as in Eq. (15) where the w controls the balance.

$$L = L_p + w L_f \tag{15}$$

D. Anomaly Sample Simulation

Since the ADRS task is unsupervised while the large margin condition mentioned above needs to be satisfied in labeled samples, we propose the simulation strategy to generate the paired anomaly samples. To simulate samples covering all the remote sensing modalities, we simulate both the anomalies in spectral domain and in spatial domain. Spectral anomalies deviate from the surroundings with properties in both spectral

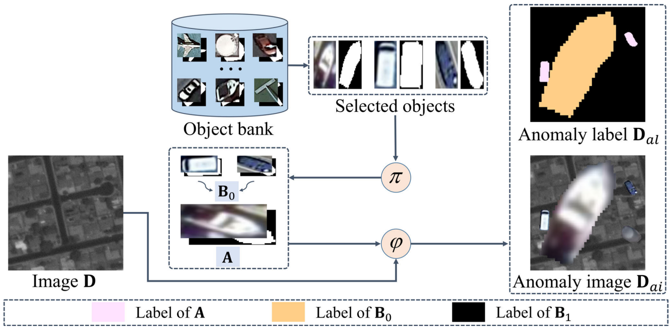


Fig. 5. The designed workflow for the spatial anomaly simulation with high spatial resolution images as input. We built an additional object bank with over 650000 instances, where the objects from different images are randomly selected and resized to preset area ranges to simulate the deviating ranking relationship of **A**, $\mathbf{B}_0$ and $\mathbf{B}_1$.

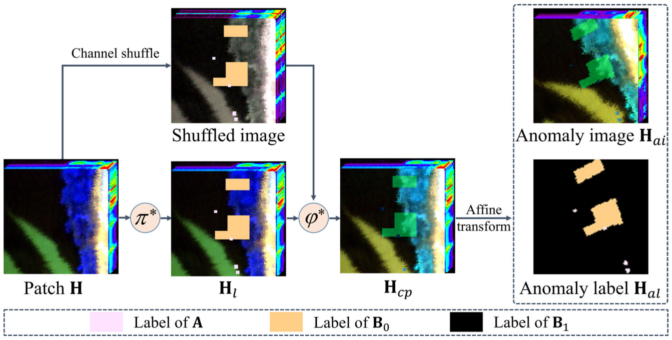


Fig. 6. The designed workflow for the spectral anomaly simulation with several hyperspectral benchmarks. We use channel shuffle operation to create the spectral deviation relationship, and the simulated anomalies have smaller size than spatial anomalies to better align with the practical situation.

Algorithm 1 UniADRS

```
Training stage (1 iteration):
    Simulate one paired spectral anomaly ( $\mathbf{H}_{ai}$ , $\mathbf{H}_{al}$ )
    Set spectral stem for $\mathbf{H}_{ai}$
    Forward computation and output one anomaly map
    Simulate one paired spatial anomaly ( $\mathbf{D}_{ai}$ , $\mathbf{D}_{al}$ )
    Set spatial stem for $\mathbf{D}_{ai}$
    Forward computation and output another anomaly map
    Compute loss $L_p$ and $L_f$ for both $\mathbf{H}_{ai}$ and $\mathbf{D}_{ai}$
    Network backward

Testing stage for any unseen image:
    Set spectral stem for hyperspectral modality or spatial stem for
      visible light, SAR, infrared and low-light modalities
    Forward computation and output one anomaly map

Model Input: One image of any remote sensing modality and scene
Model Output: One corresponding anomaly map
```

and spatial aspects (e.g., the hyperspectral modality) while the spatial anomalies deviate in the spatial properties only for the modality with few channels (e.g., SAR) [4].

For the simulation of the spatial anomalies, the workflow is designed with the large-scale iSAID dataset to provide the rich spatial details. Fig. 5 shows the overall workflow. Since the anomalies are always small in size, we simulate the large size objects $\mathbf{B}_0$ of background **B** explicitly in addition to **A** to train the model being aware of the object size. Thus, $\boldsymbol{B} = \boldsymbol{B}_0 + \mathbf{B}_1$, where $\mathbf{B}_1$ is the remaining background part. The input image $\boldsymbol{D}$ is randomly selected from the iSAID dataset, serving as the background $\mathbf{B}_1$. The anomaly tiny objects **A** and large size objects $\mathbf{B}_0$ are both selected from the pre-built object bank, which includes the 650,000 instances from the iSAID dataset. The π operation separates the selected objects into two groups (**A** and $\mathbf{B}_0$) and resizes them into the preset range (Generally, the size of $\mathbf{B}_0$ is obviously larger than **A**). Since **A** and $\mathbf{B}_0$ originally do not belong to the $\boldsymbol{D}$ and the $\mathbf{B}_0$ has an obviously larger area than **A**, the desired ranking inequality $S(\boldsymbol{a} \in \mathbf{A}) > S(\boldsymbol{b}_0 \in \mathbf{B}_0) > S(\boldsymbol{b}_1 \in \mathbf{B}_1)$ can be assumed true. Finally, the φ operation pastes the resized **A** and $\mathbf{B}_0$ into $\boldsymbol{D}$, and obtains the anomaly image $\boldsymbol{D}_{ai}$. The corresponding label $\boldsymbol{D}_{al}$ can also be obtained.

For the simulation of the spectral anomalies $(\mathbf{H}_{ai}, \mathbf{H}_{al})$, we inherit the main workflow from the prior TDD model [27], where the data argumentation technique of channel shuffling is used to create the spectral deviation of anomalies. Fig. 6 shows the simulation workflow. Given input hyperspectral patch **H**, π^* operation first randomly selects locations and obtains $\mathbf{H}_l$ for generating **A** and $\mathbf{B}_0$ according to the preset area range. The selected locations in $\mathbf{H}_l$ are then replaced by the corresponding spectra in shuffled images (i.e., φ^* operation). Since **A** and $\mathbf{B}_0$ are violently shuffled in spectral dimension and $\mathbf{B}_0$ has a larger area than **A**, the ranking $S(\boldsymbol{a} \in \mathbf{A}) > S(\boldsymbol{b}_0 \in \mathbf{B}_0) > S(\boldsymbol{b}_1 \in \mathbf{B}_1)$ can be assumed to be true in output $\mathbf{H}_{cp}$ similar to the spatial anomaly simulation. To increase the shape diversity, affine transformation is finally conducted to output the $(\mathbf{H}_{ai}, \mathbf{H}_{al})$. Three hyperspectral benchmarks (WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu) [74] are used to provide the input of the simulation workflow.

In total, the simulated anomaly samples can make the learned S be optimized with the proposed large-margin losses (Section III-C). According to the theorems proved in Section III-B, once the trained S has achieved the large-margin performance in the simulated samples, it can also detect the unseen images of the different modalities and keep the deviation inequality hold.

To show the overall workflow of proposed model, we have provided a pseudo code in Algorithm 1 including both training and testing stages.

IV. Experimental Results

A. Experimental Settings

In this section, we describe how the proposed transferring model was validated in five modalities, i.e., hyperspectral, visible light, SAR, infrared, and low-light, to show its cross-modal ability. The proposed model is named as the **uni**fied **a**nomaly **d**etector in **r**emote **s**ensing (UniADRS). In this section, the comparative experiments are firstly described with other non-transferring models, which were trained separately on each scene. Then, the model analysis results and the model efficiency are also discussed.

TABLE II

THE DETAILED INFORMATION OF CONSTRUCTED MULTI-MODAL DATASET FOR THE ADRS TASK

Modality	Source	Spatial resolution	Image size	Scene number	Anomalies
Hyperspectral	Nuance Cri; Nano-Hyperspec	4–8 cm/pixel	400×400; 200×200	82	Plastic plane, metal object, etc. [31]
Visible light	Google Earth	0.5–2 m/pixel	1044×915	100	Military camouflage [27], aircraft [12]
SAR	Gaofen-3; Sentinel-1	3–10 m/pixel	256×256	100	Various ships [10], [56]
Infrared	\	\	173×98; 407×305	100	Car, dim lamp, etc. [37]
Low-light	Indigo NV-400-M	\	2048×2048	100	Toy car, plane, tank, etc.

TABLE III

QUANTITATIVE RESULTS FOR THE HYPERSPECTRAL MODALITY, DOZENS OF SCENES IN WHU-HI PARK AND WHU-HI STATION ARE EVALUATED TOGETHER

Method	**AUC$_{(\mathrm{D,F})}$**	AUC$_{\mathrm{TD}}$	AUC$_{\mathrm{BS}}$	AUC$_{\mathrm{ODP}}$	**AUC$_{(\mathrm{D,F})}$**	AUC$_{\mathrm{TD}}$	AUC$_{\mathrm{BS}}$	AUC$_{\mathrm{ODP}}$
	Cri (1 scene)				WHU-Hi Park (27 scenes)			
GRX	0.9678	1.1932	0.8782	1.1036	0.9379	1.3091	0.8099	1.2432
ADLR	0.9579	**1.9253**	0.3159	1.2833	0.8234	1.0784	0.7995	1.2311
CRD	0.9186	1.1350	0.8738	1.0902	0.9095	1.1046	0.8514	1.1370
SC_AAE	0.8849	1.1355	0.8608	1.1114	0.9579	1.1798	0.9509	1.2149
DeepLR	0.9815	1.2465	**0.9687**	1.2337	0.9736	1.1837	**0.9607**	1.1972
TDD	0.9915	1.6298	0.8793	1.5176	0.6712	0.7391	0.0464	0.7855
UniADRS	**0.9970**	1.6755	0.9472	**1.6257**	**0.9748**	**1.4181**	0.9331	**1.3764**
	WHU-Hi Station (54 scenes)				Average			
GRX	0.8988	1.1441	0.8163	1.1628	0.9348	1.1304	0.8547	1.1146
ADLR	0.9260	1.3566	0.8650	**1.3696**	0.9024	1.0861	0.8293	1.1414
CRD	0.9722	0.9763	0.9719	1.0038	0.9334	0.9968	0.9109	1.0168
SC_AAE	0.9708	1.0455	0.9701	1.0740	0.9379	1.0611	0.9596	1.0823
DeepLR	0.9853	1.1169	**0.9825**	1.1288	0.9801	1.0914	**0.9723**	1.0999
TDD	0.7190	0.2051	0.5940	0.7991	0.7939	0.5385	0.4372	0.7519
UniADRS	**0.9860**	**1.3896**	0.9512	1.3548	**0.9859**	**1.2608**	0.9530	**1.2353**

1) Constructed Multi-Modal Dataset: We built a multi-modal dataset for the ADRS task, with hyperspectral, visible light, SAR, infrared, and low-light modalities (as detailed in Table II). The images in the dataset cover various scenes, sensor types, and resolutions. All the test images of five modalities were unseen at test stage to verify the detector transferability. The 82 hyperspectral scenes were collected from the Cri dataset [29] and the two large-scale unmanned aerial vehicle (UAV)-borne datasets of WHU-Hi-Park and WHU-Hi-Station [31]. For the low-light modality, we first captured 50 scenes at night and then doubled this by data augmentation to make the overall size balanced. The multi-modal dataset will be made publicly available.

2) Comparison Methods and Evaluation Metrics: Due to the property of the high spectral resolution, the hyperspectral modality has many unique models and was considered separately from the other modalities.

The comparative models for the hyperspectral modality were the global RX detector (GRX) [15], the abundance- and dictionary-based low-rank decomposition (ADLR) detector [75], the collaborative representation based (CRD) detector [20], the spectral constraint autoencoder (SC_AAE) detector [23], the deep low-rank prior based detector (DeepLR) [31], and the TDD method [27]. The comparison methods cover the three categories of statistical-based, representation-based, and deep learning based methods.

The comparative models for the remaining four modalities were GRX [50], a convolutional autoencoder (CAE) [76], a variational autoencoder (VAE) [77], the saliency-based method proposed by Cai et al [41] and an adversarial autoencoder (AAE) [21]. The implementation of these methods was adapted from the related ADRS studies [21], [41], [50], [77]. Besides, we also compared UniADRS with the state-of-art industrial anomaly detection model UniAD [35]. To adapt the UniAD for the small objects in ADRS task, the input size is increased from 224 to 1024. The remaining settings are kept same as [35].

The detection performance is evaluated with multi-parameter 3D receiver operating characteristic (3D ROC) curves [78]. Compared to 2D ROC curves, the threshold dimension is additionally considered and can provide more comprehensive information. The used metrics are the widely used AUC$_{(\mathrm{D,F})}$, the target detectability AUC$_{\mathrm{TD}}$, the background suppressibility AUC$_{\mathrm{BS}}$, and the overall detection probability AUC$_{\mathrm{ODP}}$. Each metric value is positively correlated with the detection performance.

3) Implementation Details: The hyperparameters of the comparative hyperspectral models are set following [27]. The CAE architecture was Unet with a ResNet50 backbone. For the SAR modality, speckle removal was conducted before applying the AAE method, following [21]. When simulating the spectral anomalies, we controlled the $\mathbf{A}$ area in ratio range [0.0064, 0.0225] and $\mathbf{B}_0$ in range [0.0225, 0.5]. Similarly, we controlled the $\mathbf{A}$ in the range [0.02, 0.06] and $\mathbf{B}_0$ in range [0.06, 0.5] for simulated spatial anomalies. The feature-level optimization loss and the pixel-level loss were added at a ratio of $w = 0.1$. The UniADRS was optimized with the Adam

TABLE IV
QUANTITATIVE RESULTS FOR THE VISIBLE LIGHT, SAR, INFRARED, AND LOW-LIGHT MODALITIES

Method	$\mathbf{AUC_{(D,F)}}$	AUC_{TD}	AUC_{BS}	AUC_{ODP}	$\mathbf{AUC_{(D,F)}}$	AUC_{TD}	AUC_{BS}	AUC_{ODP}
	(a) Visible light modality				(b) SAR modality			
GRX	0.7292	**1.1506**	0.5210	0.9425	0.8938	**1.5250**	0.7931	**1.4243**
CAE	0.7970	0.8771	0.7715	0.8516	0.8281	0.9118	0.8210	0.9047
VAE	0.6891	1.0159	0.5552	0.8819	0.8816	1.3315	0.8495	1.2995
Cai *et al.*	0.7567	0.9205	0.7005	0.8644	0.8610	1.0612	0.8347	1.0349
AAE	0.7101	0.9260	0.6375	0.8534	0.8831	0.9699	0.8757	0.9626
UniAD	0.8546	1.0217	0.7931	**0.9603**	0.9102	1.0678	0.8329	0.9905
UniADRS	**0.8948**	0.9207	**0.8901**	0.9160	**0.9595**	0.9959	**0.9549**	0.9913
	(c) Infrared modality				(d) Low-light modality			
GRX	0.6814	1.0899	0.4543	0.8629	0.6684	**1.0900**	0.4647	**0.8863**
CAE	0.8291	0.9297	0.8180	0.9187	0.6246	0.6620	0.6005	0.6380
VAE	0.7301	**1.2339**	0.4902	0.9941	0.5703	0.7299	0.4899	0.6495
Cai *et al.*	0.8853	1.2242	0.8415	**1.1805**	0.8248	0.9900	0.8049	0.9701
AAE	0.7557	1.0686	0.6598	0.9727	0.6694	0.8224	0.6196	0.7726
UniAD	0.8348	0.9145	0.8054	0.8850	0.7716	0.8563	0.7343	0.8191
UniADRS	**0.9437**	0.9820	**0.9394**	0.9778	**0.8336**	0.8558	**0.8291**	0.8513

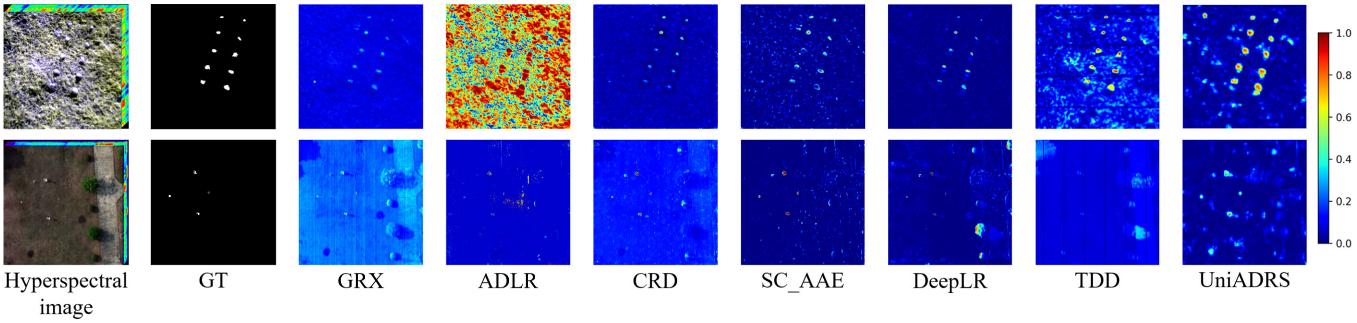


Fig. 7. Typical anomaly detection results for the hyperspectral modality, where the anomalies include rocks (first row), fabric camouflage objects (second row) and metal objects (second row).

optimizer (learning rate 0.01, weight decay 1e−5, batch size 1) over 100 epochs.

At test stage, we use the trained UniADRS on unseen images without any further fine-tuning. Spectral stem is used for hyperspectral modality and spatial stem for visible light, SAR, infrared and low-light modalities. We use the channel processing technique from [27] to deal with the varying channels of hyperspectral modality. Specifically, the image channels of various spectral datasets are interpolated into 270 at training stage, where 270 is the largest number of channels in existing anomaly detection datasets (WHU-Hi Station [31]). Interpolation operation is applied to spatial anomalies as well. Overlap technique is also used to process large images [27], Overlap technique is also used to process large images [27], where we set patch size 50 for hyperspectral modality and 100 for other modalities. All the patches are resized to the same size of 224 for unified processing. The sensitivity analysis about the inferring patch size is reported in Section IV-C.5. The CPU was an Intel(R) Xeon(R) Gold 5218R CPU @ 2.10 GHz with 251 GB memory, and the GPU was a NVIDIA GeForce RTX 4090 with 24 GB memory.

B. Comparison Results

In all the five modalities, proposed UniADRS inferred the test images directly while the comparative models were retrained for each image. The quantitative results are reported in Table III and Table IV. Fig. 7-11 visualizes the obtained anomaly maps on five modalities, respectively.

1) Hyperspectral Modality: In Table III, UniADRS is the only model that achieves an $AUC_{(D,F)}$ metric score of higher than 0.97 and an AUC_{ODP} metric score of higher than 1.35 on all three datasets (82 hyperspectral scenes).Although the TDD model shows satisfactory transferability on the Cri dataset, the metric scores drop dramatically on the UAV-borne WHU-Hi Park and Station datasets ($AUC_{(D,F)}$0.67 and 0.71, respectively). Despite the tiny anomaly sizes (especially the second example in Fig. 7), the obtained anomaly map of UniADRS has the best discriminability.

2) Visible Light Modality: Table IV(a) reports the related results. UniADRS achieves the best performance under the $AUC_{(D,F)}$ and AUC_{BS} metrics. Proposed model and UniAD are the only two models with an $AUC_{(D,F)}$ score of higher than 0.80. Although the AUC_{TD} score of our model is lower than that of GRX, this could be improved with a simple post-processing of image stretching. In Fig. 8, the first sample is inconspicuous and many detectors fail to find it. The second scene comes from the Russo-Ukrainian War, where a Russia tank is hiding and a Ukrainian UAV attempted to blow up. Many of the methods correctly locate the anomalies in this scene, but with an incomplete shape. In contrast, UniADRS achieves the best tradeoff between detection completeness and false alarms.

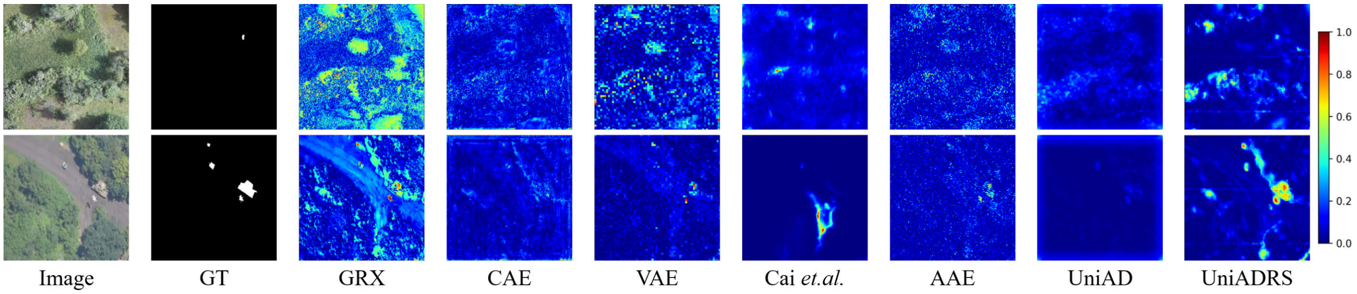


Fig. 8. Typical anomaly detection results for the visible light modality, where the anomalies include the camouflage net (first row), a tank and a drone (second row).

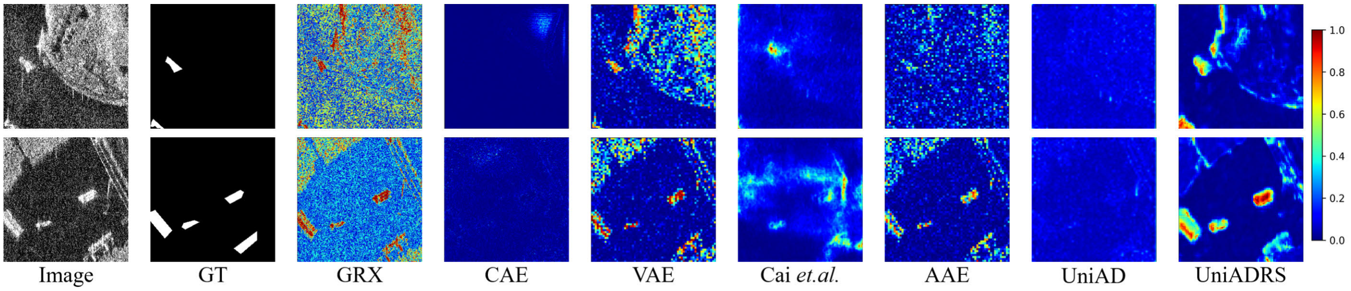


Fig. 9. Typical anomaly detection results for the SAR modality, where the anomalies include various ships.

3) SAR Modality: Table IV(b) reports the related results. The detection on SAR modality is relatively easy for most of the SAR scenes, because the anomalies (i.e., ships) lie in a homogeneous sea background. Most models can achieve the $\text{AUC}_{(\text{D,F})}$ higher than 0.85. Similar to the visible light modality, proposed model surpasses the second-place UniAD by around 4 points on $\text{AUC}_{(\text{D,F})}$ metric. For the examples in Fig. 9, many models fail to process the speckle noise and the obtained anomaly maps are full of salt-and-pepper noise such as the GRX, VAE and AAE. Since our large-margin learning has seen many spatial anomalies and learned the context modeling ability, proposed model can suppress most noises successfully.

4) Infrared Modality: Table IV(c) reports the related results. Proposed model achieves the highest $\text{AUC}_{(\text{D,F})}$ score of 0.94, which surpasses the supervised result 0.91 in [39], even when inferred directly. In Fig. 10, the anomalies in first example are extremely tiny and many model fails to detect it. The second example has 6 anomalies in total. In manual interpretation, only 2-3 anomalies can be seen in many comparative anomaly maps while our anomaly map can find 5 anomalies easily.

5) Low-Light Modality: Table IV(d) reports the related results. The captured low-light dataset seems more challenging than the remaining modalities due to the night environment. Many models achieve $\text{AUC}_{(\text{D,F})}$ lower than 0.70 while our model can still get the optimal result 0.84, showing a robust transferring ability. Due to the camouflage property of given examples in Fig. 11, proposed model is the only to locate the anomaly with discriminative boundary and high confidence.

C. Model Analysis

1) Ablation of the Model Optimization: Pixel-level and feature-level optimization are proposed for the large-margin deviation ranking target. To show the superiority, we compared it with prior ranking losses (proxy cross-entropy [71] and the average precision ranking [72]), large margin losses (sigmoid and hinge losses) [73] and the proposed pixel-level loss only. We integrate the large-margin losses into our differentiable AUC framework for fair comparison. Table V reports the related results. The results with different large-margin surrogate losses show better performance than the average precision ranking, which are consistent with our proven Theorem 1. Although cross-entropy loss is designed originally for the classification task, it has shown strong robustness for our deviation ranking task. Benefiting from considering both the ranking margin and the rationality of the threshold, proposed pixel-level loss has achieved the best transferring performance than the prior ranking losses and large-margin losses. Optimizing the model with pixel-level and feature-level losses together, the average $\text{AUC}_{(\text{D,F})}$ performance is promoted further from 0.9293 to 0.9416.

2) Statistics of the Cover Radius δ: As proven in the Theorem. 1, the cover radius δ measures the difference of simulated labeled samples and the unseen samples at the test stage, which is positively related with the demanded lowest margin in labeled samples for the transferring ability. The quantitative results in Section IV-B have already shown the model meets the lowest margin demand in simulated samples and achieve transferring ability. We analyze the δ further in this section to show the learned representative distance of different modalities.

For each pixel of unseen images, its δ is the smallest radius with the same kind of pixels (anomaly or background) in simulated samples. For each test modality, the modality δ is treated as the max δ of all the pixels (defined in Theorem 1). The radius is computed with the corresponding Euclidean distance in the feature space of $\mathbf{F}$ (defined in Section III-C). We report the resulting δ with different number of simulated images (20, 40, 60, 80), and each result is repeated four times

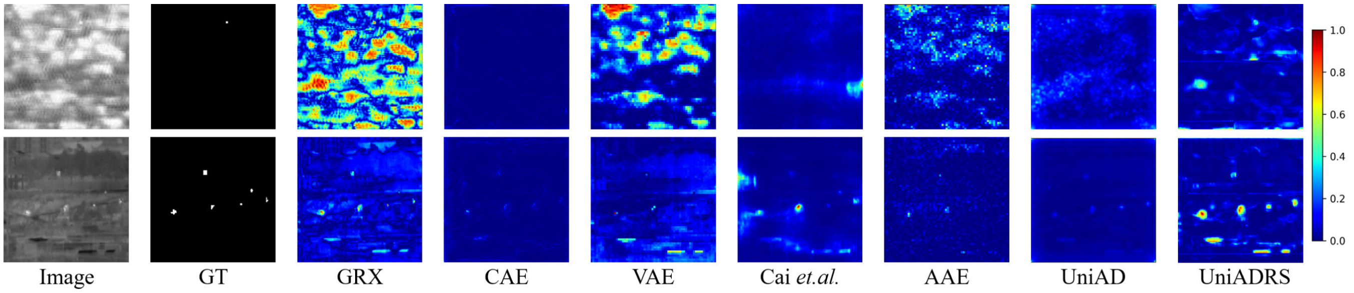


Fig. 10. Typical anomaly detection results for the infrared modality, where the anomalies include the plane (first row) and peoples (second row).

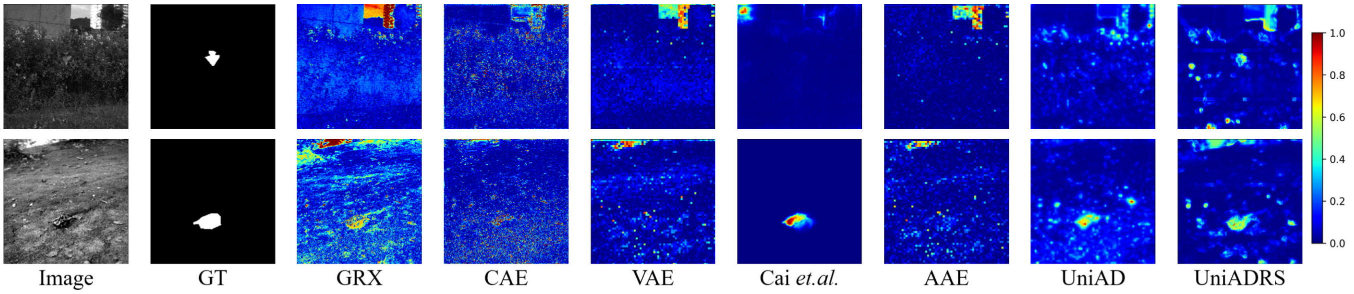


Fig. 11. Typical anomaly detection results for the low-light modality, where the anomalies include a toy plane (first row) a toy tank (second row).

TABLE V

ABLATION RESULTS FOR THE DESIGNED MODEL OPTIMIZATION LOSS

Optimization loss	Hyperspectral	Visible light	SAR	Infrared	Low-light	Average
	AUC$_{(D,F)}$	AUC$_{(D,F)}$	AUC$_{(D,F)}$	AUC$_{(D,F)}$	AUC$_{(D,F)}$	AUC$_{(D,F)}$
Proxy cross-entropy ranking	0.9409	0.8937	**0.9634**	0.9296	0.8213	0.9187
Average precision ranking	0.9252	0.8464	0.8891	0.8755	0.7194	0.8511
AUC+large margin sigmoid	0.9533	0.8896	0.9145	0.8802	0.8507	0.8977
AUC+large margin hinge	0.9308	0.8616	0.9310	0.8318	0.7924	0.8695
Pixel level	0.9641	0.8851	0.9506	0.9337	**0.8434**	0.9293
Pixel and feature level loss	**0.9859**	**0.8948**	0.9595	**0.9437**	0.8336	**0.9416**

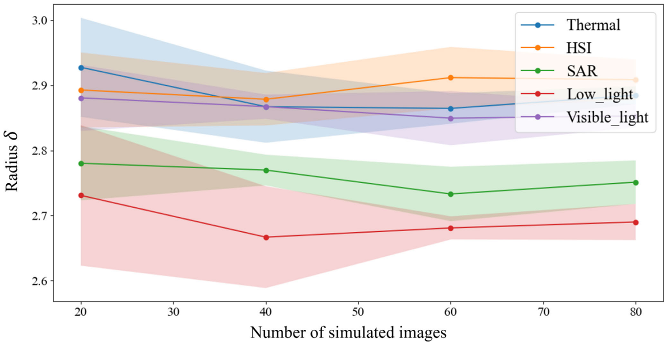


Fig. 12. The statistical radius δ between the simulated anomaly images and the unseen test images. Since the max radius value for the activated feature is in range [0, 64], the low radiuses in [2.6, 3.0] imply a low margin demand in simulated samples and the high transferring robustness.

to compute the mean (represented in broken line) and standard deviation ((represented in color block).

In the results of all the five modalities (Fig. 12), our model has obtained similar representation between the simulated images and the unseen test images, where the radius δ is lower than 3.0. Since the feature dimension of $\mathbf{F}$ is 64 in practice and the radius range is between [0, 64] after the sigmoid activation. The radius 3.0 is very small compared the max value 64. As the selected image number grows, closer representation may appear and the resulting δ decreases in many modalities (e.g., visible light and SAR). The low δ value implies the low margin demand in simulated samples and the high transferring robustness.

3) Ablation of the Sample Simulation Strategy: For the proposed UniADRS model, we simulate both the spectral anomalies and spatial anomalies, where the background, large normal objects, and anomalies are explicitly modeled. With the simulated samples, UniADRS can be trained with the designed large margin losses. The prior qualitative comparing results have shown that the explicit model for large normal objects $\mathbf{B}_0$ can decrease the false alarms effectively.

We conducted the ablation experiments from two aspects: whether to simulate the large normal objects $\mathbf{B}_0$ and whether to simulate both kinds of anomalies. The related results are shown in Table VI. It is worth noting that when there is only spectral anomaly at training stage, all test modalities are processed by the trained spectral stem, and when there is only spatial anomaly, they are all processed by the spatial stem similarly. Comparing row 1 with row 2, and row 3 with row 4, it is clear that the $\mathbf{B}_0$ simulation can bring a stable gain in most modalities especially for the hyperspectral (7 points)

TABLE VI

ABLATION RESULTS FOR THE DESIGNED ANOMALY SAMPLE SIMULATION STRATEGY

Spectral anomalies	Spatial anomalies	$\mathbf{B}_0$ simulation	Hyperspectral	Visible light	SAR	Infrared	Low-light	Average
			$\mathrm{AUC}_{(\mathrm{D,F})}$	$\mathrm{AUC}_{(\mathrm{D,F})}$	$\mathrm{AUC}_{(\mathrm{D,F})}$	$\mathrm{AUC}_{(\mathrm{D,F})}$	$\mathrm{AUC}_{(\mathrm{D,F})}$	$\mathrm{AUC}_{(\mathrm{D,F})}$
√	×	×	0.8668	0.7390	0.8674	0.7683	0.6988	0.8106
√	×	√	0.9377	0.8285	0.8038	0.8012	0.7375	0.8549
×	√	×	0.9256	0.7916	0.9211	0.8168	0.8136	0.8743
×	√	√	0.9538	0.8597	**0.9667**	0.8703	0.7815	0.9056
√	√.	√	**0.9859**	**0.8948**	0.9595	**0.9437**	**0.8336**	**0.9416**

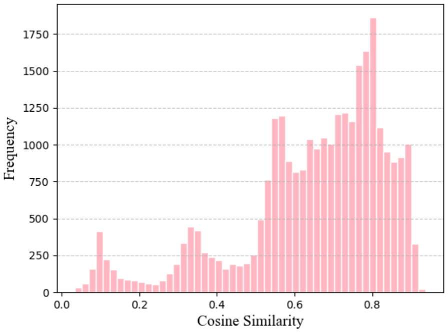


Fig. 13. The statistical cosine similarity between the simulated anomaly spectra and the background. Most spectra are in range [0.5, 0.9], which shows a weak spectra difference and high detection difficulty.

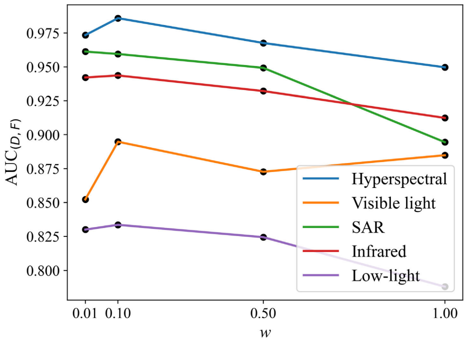


Fig. 14. The sensitivity analysis about the loss weighted parameter w, where most modalities achieve the best accuracy at the w 0.1.

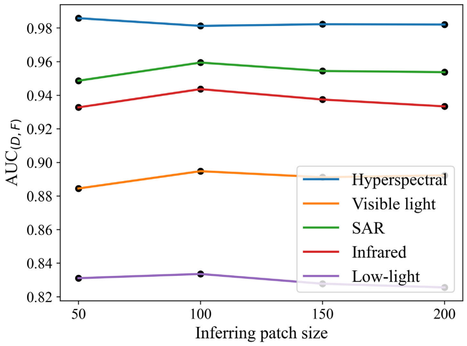


Fig. 15. The sensitivity analysis about the inferring patch size, where proposed model is robust to the changes with a maximum difference of 1 point.

and the visible light (9 points) modalities. For the infrared and low-light modalities, the increase is relatively lower (around 4 points). We deduce that the gain obtained from the $\mathbf{B}_0$ simulation is positively correlated with the scene complexity. Comparing the results of using spectral anomalies or spatial anomalies only, the spatial anomaly simulation can result in a more robust performance in most modalities, regardless of the $\mathbf{B}_0$ simulation. The inclusion of the spatial and spectral anomalies helps the detector to better fuse both the spatial and spectral features.

4) Difficulty of the Simulated Spectral Anomalies: Channel shuffling operation is used to decrease the spectral correlation of simulated anomalies and the background. Generally, the higher the correlation, the greater the detection difficulty. To quantitatively analyze the sample difficulty, we use the cosine similarity to compute the correlation degree.

Fig. 13 shows the statistical results from over 10000 spectral anomalies. For each simulated anomaly spectrum, we recorded the cosine distance between it and the surrounding background. The resulting statistical distribution is not uniform, where most results lie at the range from 0.5 to 1.0 and the peak value appears around the 0.8, implying a high correlation and detection difficulty. From this perspective, the simulated spectral samples are hard examples, which helps the learned model be more robust for the unseen anomalies.

5) Sensitivity Analyses: UniADRS is trained to be a deviation metric with designed large-margin ranking losses, where pixel-level and feature-level losses are weighted together with w to supervise the model. To report the related sensitivity analysis, we varied the w from 0.01 to 1.0 and observed the corresponding accuracy in five test modalities. The results are reported in Fig. 14. The changing of w has an obvious effect on all the modalities, which can cause a maximum difference of 6 points in accuracy. It implies that feature-level and pixel-level optimizations may not be entirely consistent and require some designed reconciliation. Except for SAR, all other modalities achieve the best accuracy at the w 0.1, which is finally chosen as the default setting.

At the test stage, we use overlap setting to improve the performance, where each image is inferred in overlapped patches. Fig. 15 reports the related sensitivity analysis about the inferring patch size. The results show that proposed model is robust to inference sizes, causing at most a 1-point difference. From the perspective of best accuracy, the optimal size for the hyperspectral modality is 50 and 100 for other modalities.

D. Model Efficiency

One of the great advantages of the proposed UniADRS model is the elimination of training for each unseen image.

TABLE VII
EFFICIENCY COMPARISON FOR THE HYPERSPECTRAL MODALITY

Method	Cri	WHU-Hi Park	WHU-Hi Station
GRX	3.73s	51.91s	71.96s
ADLR	1258.50s	25405.03s	34227.51s
CRD	1024.84s	37427.60s	37181.45s
SC_AAE	128.91s	9944.33s	21728.03s
DeepLR	31.49s	10187.13s	13714.35s
TDD	4.21s	94.51s	166.38s
UniADRS	5.14s	59.13s	121.98s

TABLE VIII
EFFICIENCY COMPARISON FOR THE VISIBLE LIGHT, SAR, INFRARED, AND LOW-LIGHT MODALITIES

Method	Visible light modality	SAR modality	Infrared modality	Low-light modality
GRX	37.52s	56.67s	41.08s	123.12s
CAE	172.64s	162.23s	646.43s	129.30s
VAE	113.77s	227.38s	71.60s	185.98s
Cai *et al.*	268.73s	371.39s	1236.66s	1452.26s
AAE	160.63s	153.19s	659.41s	1202.76s
UniAD	7440.16s	3120.53s	5760.88s	6324.86s
UniADRS	83.68s	107.69s	64.64s	61.75s

In this section, the efficiency of UniADRS is investigated by computing the model processing time for each modality.

Table VII lists the recorded processing times for the hyperspectral modality. Since the comparative models belong to transductive models and need to be trained with test images, their recorded processing times include both the training and testing stages. In contrast, proposed model can infer the unseen test modalities directly and the recorded processing time includes the testing time only. The current state-of-the-art model of DeepLR needs around 3 and 4 hours for the WHU-Hi Park and WHU-Hi Station datasets, respectively. Although TDD can deal with the WHU-Hi scenes in less than 2 min, the accuracy is not satisfactory, as shown in Table III. Keeping the highest accuracy performance, the proposed UniADRS model can process the scenes faster than the representation-based and deep learning based methods, and the time is closer to that of GRX.

Table VIII lists the recorded processing times for the remaining four modalities without spectral information. Proposed UniADRS model has surpassed all the comparative deep models by at least an order of magnitude, and achieved closer performance with GRX. Low-light modality is a special case, where GRX takes more time than proposed model due to its large image size (2048 × 2048 in Table II). Given the same image size, GRX processes the image pixel-by-pixel with CPU while proposed model can utilize the parallel computing capability of the GPU and constitute a batch for a single forward propagation.

The obtained results fully prove the real-time performance of UniADRS, and its ability to process large-scale hyperspectral scenes in real time.

V. CONCLUSION

In this study, we designed a transferring anomaly detector for different remote sensing modalities by transferring the learning target from certain image distribution to the image-independent deviation metric. To guide the learning of deviation metric, we firstly theoretically prove that although the cross-modality images are unseen at training stage, once the learned metric can rank the training samples with a large margin, it can rank the deviation score of unseen anomalies and background correctly. To satisfy the condition, we instantiate the deviation metric as a learned model and optimize it with proposed pixel-level and feature-level large-margin losses. The pixel-level loss is derived directly from the classical ranking metric AUC, where the discrete zero-one loss is replaced with the designed differentiable log loss. The feature-level loss optimizes the deviation ranking of extracted features in an equivalent way, which enlarges the distance of the enclosing hypersphere centers between the anomaly and background features. With simulated anomalies, both pixel-level and feature-level optimization work together to learn the transferring deviation metric, which is validated with five remote sensing modalities.

Focusing on the deviation learning target, this study instantiates the learnable deviation metric with a simple multi-scale convolutional network. Some potentially useful technologies such as transformer block, large-scale self-supervising are not used. Besides, feature-level and pixel-level ranking losses were found not to be completely mutually beneficial at the training stage (as in Fig. 14), implying the simple weighting method can be further improved.

UniADRS has unified the anomaly detection task for different modalities. Despite this, anomaly detection is the first step to extract the potential targets and the detectors cannot distinguish between real anomalies and detections that are not of interest. The latter recognition step is necessary for practical application [33]. To date, few studies have tried to combine the tasks and construct a complete detection and recognition pipeline. Leveraging the zero-shot anomaly detection ability of UniADRS and the zero-shot recognition ability of many foundation models to construct the complete "detection-recognition" pipeline is our next goal.

REFERENCES

[1] "Remote sensing: Rainforest shrinkage," *Nature*, vol. 454, no. 7201, p. 140, 2008. [Online]. Available: https://www.nature.com/articles/454140d#article-info

[2] S. Sun, J. Liu, X. Chen, W. Li, and H. Li, "Hyperspectral anomaly detection with tensor average rank and piecewise smoothness constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8679–8692, Nov. 2022.

[3] J. E. Fowler and Q. Du, "Anomaly detection and reconstruction from random projections," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 184–195, Jan. 2012.

[4] C.-I. Chang, S. Chen, S. Zhong, and Y. Shi, "Exploration of data scene characterization and 3D ROC evaluation for hyperspectral anomaly detection," *Remote Sens.*, vol. 16, no. 1, p. 135, Dec. 2023.

[5] W. Qiao, "Research framework of remote sensing monitoring and real-time diagnosis of Earth surface anomalies," *Acta Geodaetica Cartographica Sinica*, vol. 51, no. 7, pp. 1141–1152, 2022.

[6] J. M. Meyer, R. F. Kokaly, and E. Holley, "Hyperspectral remote sensing of white mica: A review of imaging and point-based spectrometer studies for mineral resources, with spectrometer design considerations," *Remote Sens. Environ.*, vol. 275, Jun. 2022, Art. no. 113000.

[7] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[8] S. L. Al-Khafaji, J. Zhou, X. Bai, Y. Qian, and A. W. Liew, "Spectral–spatial boundary detection in hyperspectral images," *IEEE Trans. Image Process.*, vol. 31, pp. 499–512, 2022.

[9] M. Marom, R. M. Goldstein, E. B. Thornton, and L. Shemer, "Remote sensing of ocean wave spectra by interferometric synthetic aperture radar," *Nature*, vol. 345, no. 6278, pp. 793–795, Jun. 1990.

[10] H. Guo, X. Yang, N. Wang, and X. Gao, "A CenterNet++ model for ship detection in SAR images," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107787.

[11] Z. Zheng, Y. Zhong, J. Wang, A. Ma, and L. Zhang, "FarSeg++: Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13715–13729, Nov. 2023.

[12] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.

[13] K. Jiang et al., "E2E-LIADE: End-to-end local invariant autoencoding density estimation model for anomaly target detection in hyperspectral image," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 11385–11396, Nov. 2022.

[14] J. Liu, Z. Hou, W. Li, R. Tao, D. Orlando, and H. Li, "Multipixel anomaly detection with unknown patterns for hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5557–5567, Oct. 2022.

[15] 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, Aug. 1990.

[16] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.

[17] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of Gaussian for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, Sep. 2021.

[18] Q. Ling, Y. Guo, Z. Lin, and W. An, "A constrained sparse representation model for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2358–2371, Apr. 2019.

[19] H. Su, Z. Wu, A.-X. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 195–211, Nov. 2020.

[20] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[21] M. Muzeau, C. Ren, S. Angelliaume, M. Datcu, and J.-P. Ovarlez, "SAR anomalies detection based on deep learning," in *Proc. 28th Colloque GRETSI*, 2022, pp. 1–5.

[22] H.-C. Shin and K. Na, "Anomaly detection using elevation and thermal map for security robot," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2020, pp. 1760–1762.

[23] W. Xie, J. Lei, B. Liu, Y. Li, and X. Jia, "Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection," *Neural Netw.*, vol. 119, pp. 222–234, Nov. 2019.

[24] X. Lu, W. Zhang, and J. Huang, "Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1527–1537, Mar. 2020.

[25] P. Sprechmann, A. M. Bronstein, and G. Sapiro, "Learning efficient sparse and low rank models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1821–1833, Sep. 2015.

[26] N. Huyan, X. Zhang, D. Quan, J. Chanussot, and L. Jiao, "AUD-Net: A unified deep detector for multiple hyperspectral image anomaly detection via relation and few-shot learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 5, pp. 6835–6849, May 2024.

[27] J. Li, X. Wang, S. Wang, H. Zhao, L. Zhang, and Y. Zhong, "One-step detection paradigm for hyperspectral anomaly detection via spectral deviation relationship learning," 2023, *arXiv:2303.12342*.

[28] I. Ahmed, T. Galoppo, X. Hu, and Y. Ding, "Graph regularized autoencoder and its application in unsupervised anomaly detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4110–4124, Aug. 2022.

[29] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503314.

[30] J. Li, X. Wang, H. Zhao, S. Wang, and Y. Zhong, "Anomaly segmentation for high-resolution remote sensing images based on pixel descriptors," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 4, pp. 4426–4434.

[31] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Deep low-rank prior for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527017.

[32] C. Lin, S.-Y. Chen, C.-C. Chen, and C.-H. Tai, "Detecting newly grown tree leaves from unmanned-aerial-vehicle images using hyperspectral target detection techniques," *ISPRS J. Photogramm. Remote Sens.*, vol. 142, pp. 174–189, Aug. 2018.

[33] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.

[34] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.

[35] Z. You et al., "A unified model for multi-class anomaly detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, Dec. 2022, pp. 4571–4584.

[36] Y. Gong, X. Yu, Y. Ding, X. Peng, J. Zhao, and Z. Han, "Effective fusion factor in FPN for tiny object detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1159–1167.

[37] H. Wang, L. Zhou, and L. Wang, "Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8508–8517.

[38] S.-Y. Chen, Y. Wang, C.-C. Wu, C. Liu, and C.-I. Chang, "Real-time causal processing of anomaly detection for hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 1511–1534, Apr. 2014.

[39] S. Sun, J. Liu, and W. Li, "Spatial invariant tensor self-representation model for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 54, no. 5, pp. 3120–3131, May 2024.

[40] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 597–601, May 2017.

[41] C. Lile and L. Yiqun, "Anomaly detection in thermal images using deep neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2299–2303.

[42] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.

[43] J. Lei, W. Xie, J. Yang, Y. Li, and C.-I. Chang, "Spectral–spatial feature extraction for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8131–8143, Oct. 2019.

[44] C.-I. Chang, "Target-to-anomaly conversion for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5540428.

[45] C.-I. Chang, "Constrained energy minimization anomaly detection for hyperspectral imagery via dummy variable trick," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517119.

[46] T. P. Watson et al., "Evaluation of aerial real-time RX anomaly detection," *Proc. SPIE*, vol. 12519, pp. 254–260, Jun. 2023.

[47] S. Liu et al., "Hyperspectral real-time online processing local anomaly detection via multiline multiband progressing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5518719.

[48] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005.

[49] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.

[50] Y. Haitman, I. Berkovich, S. Havivi, S. Maman, D. G. Blumberg, and S. R. Rotman, "Machine learning for detecting anomalies in SAR data," in *Proc. IEEE Int. Conf. Microw., Antennas, Commun. Electron. Syst. (COMCAS)*, Nov. 2019, pp. 1–5.

[51] P. K. Pokala, R. V. Hemadri, and C. S. Seelamantula, "Iteratively reweighted minimax-concave penalty minimization for accurate low-rank plus sparse matrix decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8992–9010, Dec. 2022.

[52] W. Sun, C. Liu, J. Li, Y. M. Lai, and W. Li, "Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 8, no. 1, May 2014, Art. no. 083641.

[53] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[54] Y. Zhang, D. Bo, L. Zhang, and S. Wang, “A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2015.

[55] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, “Anomaly detection in hyperspectral images based on low-rank and sparse representation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.

[56] N. Wang, B. Li, Q. Xu, and Y. Wang, “Automatic ship detection in optical remote sensing images based on anomaly detection and SPP-PCANet,” *Remote Sens.*, vol. 11, no. 1, p. 47, Dec. 2018.

[57] T. Jiang, W. Xie, Y. Li, J. Lei, and Q. Du, “Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6504–6517, Nov. 2022.

[58] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, “GAN-based hyperspectral anomaly detection,” in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 1891–1895.

[59] T. Jiang, W. Xie, Y. Li, and Q. Du, “Discriminative semi-supervised generative adversarial network for hyperspectral anomaly detection,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 2420–2423.

[60] M. Díaz, R. Guerra, P. Horstrand, S. López, and R. Sarmiento, “A line-by-line fast anomaly detector for hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8968–8982, Nov. 2019.

[61] J. López-Fandiño, D. B. Heras, and F. Argüello, “Using heterogeneous computing and edge computing to accelerate anomaly detection in remotely sensed multispectral images,” *J. Supercomput.*, vol. 80, no. 9, pp. 12543–12563, Jun. 2024.

[62] A. Ruhan, X. Mu, L. Feng, and J. He, “A fast recursive LRX algorithm with extended morphology profile for hyperspectral anomaly detection,” *Can. J. Remote Sens.*, vol. 47, no. 5, pp. 731–748, Sep. 2021.

[63] M. Coca and M. Datcu, “FPGA accelerator for meta-recognition anomaly detection: Case of burned area detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1–13, 2023.

[64] H. Su, Z. Wu, H. Zhang, and Q. Du, “Hyperspectral anomaly detection: A survey,” *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022.

[65] R. Zhao, B. Du, L. Zhang, and L. Zhang, “A robust background regression based score estimation algorithm for hyperspectral anomaly detection,” *ISPRS J. Photogramm. Remote Sens.*, vol. 122, pp. 126–144, Dec. 2016.

[66] C.-I. Chang, “An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis,” *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.

[67] L. S. Kalman and E. M. Bassett III, “Classification and material identification in an urban environment using HYDICE hyperspectral data,” *Proc. SPIE*, vol. 3118, pp. 57–68, Oct. 1997.

[68] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, “Development of a web-based application to evaluate target finding algorithms,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2008, pp. 915–918.

[69] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, “MUUFL gulfport hyperspectral and LiDAR airborne data set,” Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570, 2013. [Online]. Available: https://github.com/GatorSense/MUUFLGulfport?tab=readme-ov-file

[70] T. Zhou, D. Tao, and X. Wu, “Manifold elastic net: A unified framework for sparse dimension reduction,” *Data Mining Knowl. Discovery*, vol. 22, no. 3, pp. 340–371, May 2011.

[71] S. Bruch, “An alternative cross entropy loss for learning-to-rank,” in *Proc. Web Conf.*, Apr. 2021, pp. 118–126.

[72] P. Mohapatra, M. Rolinek, C. V. Jawahar, V. Kolmogorov, and M. P. Kumar, “Efficient optimization for rank-based loss functions,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3693–3701.

[73] X.-L. Zhang and M. Xu, “AUC optimization for deep learning-based voice activity detection,” *EURASIP J. Audio, Speech, Music Process.*, vol. 2022, no. 1, pp. 1–12, Oct. 2022.

[74] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, “WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H^2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF,” *Remote Sens. Environ.*, vol. 250, Dec. 2020, Art. no. 112012.

[75] Y. Qu et al., “Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.

[76] S. Mabu, K. Fujita, and T. Kuremoto, “Disaster area detection from synthetic aperture radar images using convolutional autoencoder and one-class SVM,” *J. Robot. Netw. Artif. Life*, vol. 6, no. 1, pp. 48–51, 2019.

[77] S. Sinha et al., “Variational autoencoder anomaly-detection of avalanche deposits in satellite SAR imagery,” in *Proc. 10th Int. Conf. Climate Informat.*, Sep. 2020, pp. 113–119.

[78] C. Chang, “Multiparameter receiver operating characteristic analysis for signal detection and classification,” *IEEE Sensors J.*, vol. 10, no. 3, pp. 423–442, Mar. 2010.

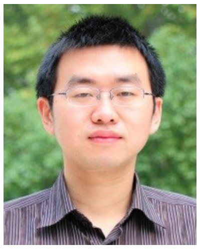
Jingtao Li received the B.S. degree from the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China, in 2021. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University.

His major research interests include anomaly detection in remote sensing.

Xinyu Wang (Member, IEEE) received the B.S. degree in photogrammetry and remote sensing and the Ph.D. degree in communication and information systems from Wuhan University, China, in 2014 and 2019, respectively.

Since 2019, he has been an Associate Research Fellow with the School of Remote Sensing and Information Engineering, Wuhan University. His major research interests include hyperspectral data processing and applications.

Hengwei Zhao (Member, IEEE) received the B.S. degree in surveying and mapping engineering from the School of Resources and Civil Engineering, Northeastern University, Shenyang, China, in 2019. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan.



Yanfei Zhong (Senior Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2002 and 2007, respectively.

Since 2010, he has been a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. He organized the Intelligent Data Extraction, Analysis and Applications of Remote Sensing (RSIDEA) Research Group. He has published more than 100 research articles in international journals, such as *Remote Sensing of Environment*, *ISPRS Journal of Photogrammetry and Remote Sensing*, and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. His research interests include hyperspectral remote sensing information processing, high-resolution remote sensing image understanding, and geoscience interpretation for multisource remote sensing data and applications.

Dr. Zhong is a fellow of the Institution of Engineering and Technology (IET). He was a recipient of the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). He won the Second-Place Prize of the 2013 IEEE GRSS Data Fusion Contest and the Single-View Semantic 3D Challenge of the 2019 IEEE GRSS Data Fusion Contest, respectively. He is currently serving as an Associate Editor for *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* and *International Journal of Remote Sensing*.