

Article

Lightweight Dual-Branch Target-Aware Network for Land-Based Hyperspectral Camouflage Target Detection

Xingxin Song ¹ , Bing Zhou ^{1,*} , Jiaju Ying ¹, Jiale Zhao ¹, Lei Deng ² and Qi Chen ¹

¹ Shijiazhuang Campus, Army Engineering University of PLA, Shijiazhuang 050003, China; songxingxin1215@163.com (X.S.); yjioptics25@126.com (J.Y.); zjl2025@aeu.edu.cn (J.Z.); cqoptics@163.com (Q.C.)

² Department of Information Engineering, Army Arms University of PLA, Hefei 230031, China; denglei144@163.com

* Correspondence: zhbngxgc@aeu.edu.cn

Highlights

What are the main findings?

- The presented lightweight dual-branch target-aware network achieved F1 scores of 0.9438–0.9601, PR-AUC values of 0.9814–0.9938, and false-alarm rates of 0.0002–0.0003 across four land-based hyperspectral camouflage datasets, outperforming eight representative detectors.
- Combining raw-spectral and whitened-PCA branches with target-aware spatial gating and center-aware classification improved robustness under severe class imbalance; with only 0.5% target pixels for fine-tuning, cross-dataset PR-AUC remained between 0.8833 and 0.9288.

What are the implications of the main findings?

- A total of 0.0556 M parameters and 8.0699 M FLOPs per patch indicate a favorable accuracy–efficiency trade-off for label-scarce hyperspectral camouflage target detection in resource-aware remote sensing applications.
- The results support cross-scene reuse of the learned spectral–spatial representation with sparse annotations, while further validation on larger multi-sensor datasets and faster full-image inference are needed for robust real-time deployment.

Abstract

Hyperspectral camouflaged target detection is challenging because target and background spectra are highly coupled, positive pixels are extremely scarce, and scene-dependent spectral variation limits cross-dataset robustness. Existing detectors often rely on a single spectral representation or use spatial models with substantial parameters and computation, which may discard complementary evidence or increase deployment costs. This article presents a lightweight dual-branch target-aware network for land-based hyperspectral camouflage target detection. The model preserves fine raw spectral details and whitened PCA features while integrating multi-scale spatial context. A target-aware spatial gate is further employed to suppress background-like responses, and a center-aware decision head alleviates the dilution of small targets during patch pooling. Class-weighted loss and balanced sampling address severe class imbalance. On four land-based hyperspectral camouflage datasets with 1% training, 4% validation, and 95% testing pixels, the proposed method achieves F1 scores of 0.9438–0.9601, PR-AUC values of 0.9814–0.9938, and false-alarm rates of 0.0002–0.0003, outperforming eight representative detectors. Ablation studies confirm the complementarity of the raw and PCA branches and the effectiveness of the target-aware



Academic Editors: Shou Feng, Jiaqing Qiao and Yuxiang Zhang

Received: 8 September 2026

Revised: 14 October 2026

Accepted: 15 October 2026

Published: 17 October 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

gate and center-aware head. In cross-dataset evaluation with only 0.5% target pixels for fine-tuning and another 0.5% for validation, PR-AUC remains between 0.8833 and 0.9288. The model requires only 0.0556 M parameters and 8.0699 M FLOPs per patch, indicating a favorable trade-off between accuracy, generalization, and lightweight deployment.

Keywords: hyperspectral images; camouflage target detection; lightweight deep learning; class imbalance

1. Introduction

Hyperspectral images are capable of capturing nanoscale spectral signatures of ground materials, enabling the differentiation of materials that appear nearly identical in visible-band imagery [1–3]. A land-based hyperspectral camouflaged target is defined as a target in a terrestrial scene whose material, coating, or covering is deliberately designed to mimic the spectral and spatial characteristics of the surrounding background, such as vegetation, soil, or artificial clutter. Because its spectral signature is highly similar to that of the background, it is difficult to distinguish using conventional visible-band or broadband imagery. Land-based hyperspectral camouflage target detection aims to identify and localize such targets at the pixel or subpixel level by exploiting the fine spectral and spatial information provided by hyperspectral images. This task is of great importance in military reconnaissance, battlefield surveillance, camouflage effectiveness assessment, remote sensing interpretation, material identification, security monitoring, and environmental monitoring [4,5]. However, it is substantially more challenging than conventional image segmentation, as a detector must suppress a wide variety of background pixels while accurately identifying a very small number of target pixels [6–8].

Three factors make land-based camouflage target detection especially challenging. First, camouflage targets and surrounding backgrounds are often strongly coupled in the spectral domain. In the four datasets used in this study, the target–background imbalance ranges from 1:169 to 1:251, and the mean spectral angle distances are only 0.053 to 0.434 rad, indicating highly similar spectral signatures. Second, terrestrial backgrounds are spatially heterogeneous; jungle, bush, and grass scenes introduce different background distributions and hard-negative samples. Third, only a small proportion of pixels can be labeled in practical applications, which aggravates class imbalance and makes pixel accuracy unsuitable as the main evaluation metric [9–11].

Classical hyperspectral target detectors, including constrained energy minimization (CEM), orthogonal subspace projection (OSP), adaptive cosine/coherence estimation (ACE), and its variants, rely on target signatures or statistical background models. These methods are computationally efficient and interpretable, but their linear decision forms often become unstable when target and background spectra overlap strongly [12–15]. Sparse-representation and tensor-based detectors further exploit structural priors to separate targets from backgrounds [16]. Although such methods improve robustness in some scenes, their performance still depends heavily on the separability of the constructed representation. More recently, deep models such as Unbalanced Episode Meta-Learning (UEML), Triplet Spectralwise Transformer-based Target Detector (TSTTD), and HTD-Mamba have introduced learned spectral–spatial representations and achieved stronger nonlinear modeling capacity [17–19]. However, deep detectors often require considerable parameters or computation, and patch-level pooling can dilute small-target evidence when background pixels dominate the receptive field [20–22].

These observations reveal three coupled requirements: preserving complementary spectral evidence, improving target-aware spatial decisions under severe imbalance, and

maintaining a lightweight architecture suitable for deployment. Existing studies often address only one of these requirements. For example, PCA-based representations reduce spectral redundancy but may discard subtle band-specific differences, whereas raw spectral branches preserve such details but are more sensitive to correlated bands. Similarly, larger spatial receptive fields improve context modeling but increase computation and may introduce more background interference [23–25].

To address these issues, we propose a lightweight dual-branch target-aware network for land-based hyperspectral camouflage target detection. The model contains a raw-spectral branch and a PCA-based branch. The raw branch preserves fine spectral evidence, while the PCA branch provides decorrelated and whitened low-dimensional features. A target-aware spatial gate suppresses background-like responses after feature fusion, and a center-aware decision head combines pooled patch context with the feature at the candidate pixel. In addition, inverse-frequency class weights and balanced random sampling are used to reduce the bias caused by extremely scarce target pixels.

The main contributions of this work are summarized as follows.

1. A lightweight raw-PCA dual-branch architecture is proposed for hyperspectral camouflage target detection. The two branches preserve complementary original and globally transformed spectral evidence, while depthwise multi-scale fusion enlarges the receptive field with limited computation.
2. A target-aware spatial gate and a center-aware decision head are integrated to address background interference and small-target dilution. The gate attenuates background-like spatial responses, whereas the center branch preserves pixel-specific evidence during patch pooling.
3. A class-imbalance-aware training strategy is designed by combining weighted cross-entropy loss with balanced target/background sampling. This strategy improves detection reliability under target–background imbalance as severe as 1:251.
4. Comprehensive experiments are conducted on four land-based HSI datasets under a 1% training, 4% validation, and 95% testing protocol. The proposed method is compared with eight representative detectors and further evaluated through module ablation, cross-dataset fine-tuning, and complexity analysis. It achieves F1 scores of 0.9438 to 0.9601 and PR-AUC values of 0.9814 to 0.9938 while requiring only 0.0556 M parameters and 8.0699 M FLOPs per patch.

The remainder of this article is organized as follows. Section 2 reviews related hyperspectral target detection methods. Section 3 presents the proposed model. Section 4 describes the datasets, experimental protocol, comparison results, ablation studies, cross-dataset evaluation, and efficiency analysis. Section 5 concludes this article.

2. Related Work

2.1. Classical Hyperspectral Target Detection

Classical hyperspectral target detection methods are mainly based on matched filtering, subspace projection, and statistical hypothesis testing. CEM designs a filter that minimizes the background response while constraining the response to a known target signature. OSP projects the data onto a subspace orthogonal to the background representation, thereby suppressing known background components. ACE measures the matched angle or coherence between an observed spectrum and the target subspace, while S-ACE improves the adaptive detection formulation under varying background conditions. These methods are widely used because they are efficient, interpretable, and suitable for scenes with reliable target signatures or background statistics [26–28].

However, their decision processes are largely linear or statistically constrained. When camouflaged targets and terrestrial backgrounds share similar spectral signatures, the target signature is no longer well separated from the background distribution. Mixed pixels, illumination variation, and a spatially changing background further weaken the assumptions behind fixed projection or filtering models. As a result, classical detectors often produce either missed detections or numerous false alarms in strongly coupled camouflage scenes [29–31]. Unlike these methods, the proposed model learns a nonlinear spectral–spatial representation from scarce labeled pixels and combines raw and PCA evidence rather than relying on a single linear decision statistic.

2.2. Sparse Representation and Tensor-Based Detection

To exploit structural differences between targets and backgrounds, sparse-representation and tensor-based methods model hyperspectral images as combinations of low-rank background components and sparse target components. By representing image cubes or spectral vectors with structured dictionaries, these methods can improve robustness to spectral correlation and local background variation. Tensor formulations further preserve the spatial–spectral structure of HSI data instead of treating pixels independently [32,33].

Nevertheless, their performance depends on whether the imposed sparsity, low-rank, or tensor priors match the observed scene. Camouflaged targets are often small, spectrally similar to the background, and embedded in heterogeneous terrestrial clutter; under these conditions, the target component may not be sufficiently sparse or distinguishable from hard-negative background structures. In addition, parameter selection and threshold determination can be sensitive when only a few target pixels are available [34,35]. The proposed model avoids relying on an explicit low-rank/sparse decomposition. Instead, it uses a learnable dual-branch representation and target-aware gating to suppress ambiguous background responses.

2.3. Deep Spectral–Spatial Detection

Deep learning methods have been increasingly applied to hyperspectral target detection because they can learn nonlinear spectral–spatial features directly from data [36–38]. Representative models include UEML, TSTTD, and HTD-Mamba. UEML improves feature learning for target–background separation, TSTTD models spectral–spatial interactions for target detection, and HTD-Mamba introduces state-space sequence modeling into hyperspectral target detection. These methods generally achieve stronger representation capacity than classical filtering or projection-based detectors. Beyond discriminative detection, generative models such as diffusion models have recently shown strong potential in hyperspectral image super-resolution, suggesting new directions for background modeling and sample augmentation in future target detection research [39].

Despite their advantages, deep detectors face two practical difficulties in land-based camouflage scenes. First, model capacity and spatial context modeling often increase parameters and computation. Under the same test environment used in this study, UEML, TSTTD, and HTD-Mamba require 0.8226 M, 0.6726 M, and 0.3247 M parameters, respectively, and 12.8469 M, 124.7253 M, and 14.0517 M FLOPs per patch. Second, many patch-based detectors aggregate the whole patch without explicitly distinguishing the central candidate pixel from background pixels. When a target occupies only a small part of the patch, global pooling may dilute target evidence, while a larger receptive field may introduce an additional hard-negative background [40–42].

Dual-branch and multi-branch architectures have been explored in hyperspectral image processing to combine complementary features from different representations. For example, two-stream networks have been designed to jointly exploit spectral and spatial

information for classification and detection. Hybrid CNN–Transformer dual-branch models further integrate local convolutional features with global attention for improved representation. These studies show that dual-branch designs can effectively fuse heterogeneous evidence and improve robustness. However, most existing dual-branch networks are developed for classification tasks or rely on heavy backbones, and their potential for lightweight camouflage target detection under severe class imbalance remains underexplored [43–45].

Principal component analysis (PCA) has also been widely integrated into deep hyperspectral networks. In many works, PCA is used as a preprocessing step to reduce spectral redundancy and provide compact input to CNNs or transformers. Some studies further combine PCA features with raw spectral bands in a multi-branch manner to preserve both decorrelated global information and fine band-specific details. These works confirm the value of PCA in deep HSI processing. Nevertheless, most of them focus on classification or reconstruction, and the joint use of raw spectra and whitened PCA features for target detection, especially under extreme class imbalance, has not been fully investigated [46–48].

2.4. Positioning of the Proposed Method

The proposed model differs from the above methods in three aspects. First, it jointly preserves raw spectral details and whitened PCA features. PCA reduces spectral redundancy, while the raw branch retains band-specific evidence that may be compressed or removed by dimensionality reduction. Second, it introduces a target-aware spatial gate after feature fusion, allowing the detector to suppress background-like locations before decision aggregation. Third, a center-aware decision head explicitly preserves the feature of the candidate pixel, alleviating the small-target dilution caused by patch-level pooling. Combined with weighted loss and balanced sampling, the proposed network aims to achieve a practical balance among detection accuracy, false-alarm suppression, cross-scene robustness, and computational cost.

3. Method

3.1. Overall Architecture

As illustrated in Figure 1, the proposed model consists of four parts: a raw-spectral branch, a PCA branch, a target-aware spatial gate, and a center-aware decision head. The two spectral branches extract complementary representations; the gate adaptively suppresses ambiguous background responses; and the decision head combines global patch context with explicit center evidence for binary prediction.

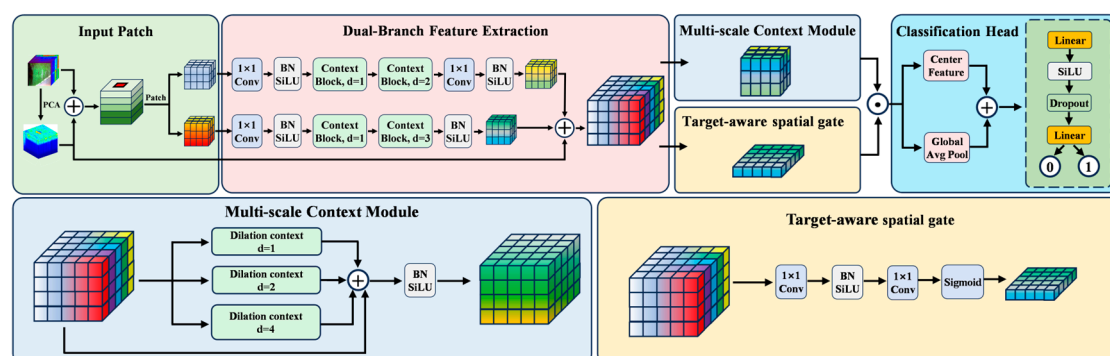


Figure 1. Overall architecture of the proposed method.

The raw branch maps the original spectral patch through a 1×1 convolution from B to 64 channels, batch normalization, and SiLU activation. Two residual context blocks with dilation rates of 1 and 2 then model local and mid-range spatial dependencies. A final 1×1 convolution projects the feature to K channels, giving

$$\mathbf{F}_r = R(\mathbf{X}_i) \in \mathbb{R}^{K \times s \times s} \quad (1)$$

The PCA branch applies a 1×1 convolution, batch normalization, and SiLU activation to the PCA patch, followed by residual context blocks with dilation rates of 1 and 3:

$$\mathbf{F}_p = P(\hat{\mathbf{X}}_i) \in \mathbb{R}^{K \times s \times s} \quad (2)$$

The two learned features and the original PCA patch are concatenated as

$$\mathbf{F}_c = [\mathbf{F}_r; \mathbf{F}_p; \hat{\mathbf{X}}_i] \in \mathbb{R}^{3K \times s \times s} \quad (3)$$

Thus, subsequent decisions are based on both learned raw-spectral evidence and globally transformed PCA evidence.

3.2. Target-Aware Spatial Gate

Camouflaged targets and surrounding background often have similar spectra. A purely global representation may therefore mix weak target evidence with strong hard-negative background evidence. To address this issue, we introduce a target-aware spatial gate that estimates, for each spatial location, how target-like the fused evidence is.

The gate is implemented as

$$\mathbf{g} = \sigma(\mathbf{W}_2 \text{SiLU}(\text{BN}(\mathbf{W}_1 \mathbf{F}_c))) \quad (4)$$

where $\mathbf{W}_1$ maps $3K$ channels to 16 channels, $\mathbf{W}_2$ maps 16 channels to one channel, and σ is the sigmoid function. The resulting $\mathbf{g} \in (0,1)^{1 \times s \times s}$ is a spatial confidence map. Values close to one indicate target-like evidence, whereas values close to zero indicate background-like evidence.

The gate is applied after multi-scale feature fusion rather than before it. This design allows the detector to first integrate raw, PCA, and dilated-context evidence and then suppress responses that remain background-like after fusion. Formally,

$$\tilde{\mathbf{F}} = \mathbf{F}_f \odot \mathbf{g} \quad (5)$$

where $\odot$ denotes broadcast element-wise multiplication across channels. The gate does not force a hard background decision; instead, it provides differentiable spatial attenuation before pooling. This mechanism is especially relevant for camouflaged HSI targets, where background and target spectra overlap, and hard-negative context can dominate pooled evidence.

3.3. Multi-Scale Context Fusion

To represent targets of different spatial extents, three depthwise-separable convolutions with dilation rates of 1, 2, and 4 are applied in parallel to $\mathbf{F}_c$. Their outputs, together with $\mathbf{F}_r$, $\mathbf{F}_p$ and $\hat{\mathbf{X}}_i$, are concatenated and projected by a 1×1 convolution to K channels:

$$\mathbf{F}_f = \phi\left(\text{BN}\left(\mathbf{W}_f[D_1(\mathbf{F}_c); D_2(\mathbf{F}_c); D_4(\mathbf{F}_c); \mathbf{F}_r; \mathbf{F}_p; \hat{\mathbf{X}}_i]\right)\right) \quad (6)$$

where ϕ denotes SiLU activation. The identity paths preserve direct spectral evidence, while the dilated paths enlarge the effective receptive field without substantially increasing parameters. For $B = 89$, $K = 32$, and $s = 9$, the detector contains 0.0556 M parameters and requires 8.0699 M FLOPs, equivalent to 4,230,960 MACs, per patch.

3.4. Center-Aware Evidence and Classification Head

Global average pooling summarizes the entire gated patch, but a small target occupies only a small fraction of a 9×9 patch. If target evidence is averaged with many background locations, the resulting vector may be diluted. Therefore, the detector explicitly preserves the center feature.

Because each patch is centered on the candidate pixel, the center location is the most direct evidence for that pixel. We extract

$$\mathbf{z}_c = \tilde{\mathbf{F}}\left(:, :, \frac{s}{2}, \frac{s}{2}\right) \quad (7)$$

to combine it with the pooled representation

$$\mathbf{z}_g = \text{GAP}(\tilde{\mathbf{F}}) \quad (8)$$

The final feature vector is

$$\mathbf{z} = [\mathbf{z}_g; \mathbf{z}_c] \in \mathbb{R}^{2K} \quad (9)$$

It is passed through layer normalization, a linear layer from $2K$ to K , SiLU activation, dropout with a probability of 0.1, and a final linear classifier. The output is

$$\hat{\mathbf{y}} = f_{\theta}(\mathbf{z}) \in \mathbb{R}^2 \quad (10)$$

The pooled vector provides patch-level context, while the center vector preserves local spectral-spatial evidence at the candidate location. Their concatenation gives the classifier both global and target-centered information.

3.5. Training and Inference

Training and validation pixels are sampled in a class-stratified manner. The loss is weighted cross-entropy, with inverse class-frequency weights normalized to sum to two. In addition, a balanced random sampler gives target and background samples equal sampling probability per epoch. Training patches are randomly flipped or rotated, while validation and test patches are not augmented. The decision threshold is selected on the validation set by maximizing validation F1 and is then applied unchanged to the test set. During full-image inference, every pixel is treated as a candidate center. Reflect padding is used to form boundary patches, and the target probability is obtained from the target-class softmax output. A binary prediction is generated by thresholding the resulting probability map. All experiments were repeated with five random seeds, and the results are reported as the mean $\pm$ standard deviation.

4. Experimental Results and Analysis

4.1. Datasets

To comprehensively evaluate the detection performance of the proposed model under complex terrestrial backgrounds, we acquired four land-based hyperspectral camouflage target datasets (Data A–D) using a tripod-mounted HSI-300 spectrometer (Gooch & Housego, Orlando, FL, USA) at 14:00 on 14 April 2026 in a representative terrestrial area of Shijiazhuang, Hebei Province, China. The detailed parameters of each dataset are summarized in Table 1, while the corresponding RGB composite images, pixel-level ground-truth masks (which were meticulously annotated using the ENVI software, version 5.6), mean spectral curves of both targets and backgrounds, along with their Spectral Angle Distance (SAD) values are illustrated in Figure 2. These quantitative SAD results explicitly characterize the high spectral coupling between targets and backgrounds, which not only

substantiates the intrinsic difficulty of the camouflage target detection task but also lays a solid data foundation for rigorously validating the proposed model's robustness and discriminative capability in distinguishing subtle spectral discrepancies [49].

Table 1. Dataset details.

	Data A	Data B	Data C	Data D
Scene type	Jungle	Jungle	Bushes	Grass
Source	HSI-300	HSI-300	HSI-300	HSI-300
Spatial size	500×500	500×500	496×600	500×500
Bands	89	89	89	89
Spectral range	449–801	449–801	449–801	449–801
Target type	Armored vehicle	Camouflage clothes	Armored vehicle	Armored vehicle
Target pixels	1466	1248	1357	993
Background pixels	248,534	248,752	296,243	249,007
Imbalance ratio	1:169	1:199	1:218	1:251

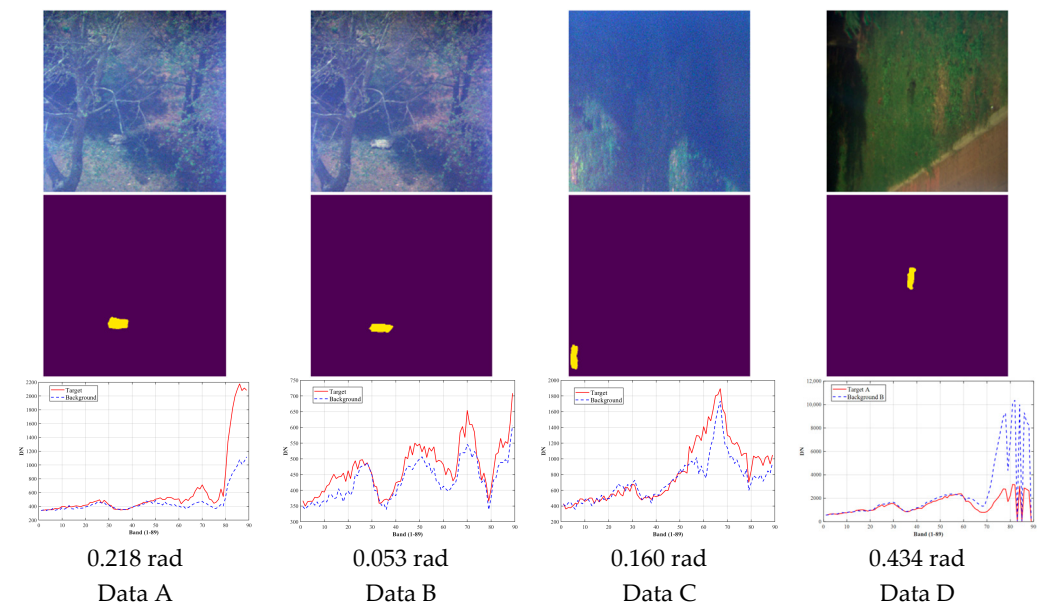


Figure 2. False-color images, ground-truth masks, spectral curves and spectral angle distance (SAD) of target-background pairs for four hyperspectral datasets.

4.2. Experimental Setup

4.2.1. Experimental Protocol

For same-dataset detection, each scene is split in a class-stratified manner into 1% pixels for training, 4% for validation, and 95% for testing. Spectral values are first min-max normalized, and band-wise mean centering is then applied. Both operations use statistics computed only from the training pixels, and the resulting parameters are fixed for validation and testing. For models using PCA features, PCA with whitening is fitted only on the training pixels of the source domain and then applied unchanged to validation and test pixels. This prevents distribution information from validation or test data from leaking into feature extraction.

The proposed model is trained with inverse-frequency class weights and balanced random sampling. Training patches are augmented by random flips and rotations, whereas validation and test patches are not augmented. The validation set is used for model selection, threshold selection, and hyperparameter analysis; test pixels are used only for

final evaluation. The binary decision threshold is obtained by maximizing F1 on the validation set and is then fixed for testing.

For cross-dataset evaluation, the model is first trained on the source dataset and evaluated on the three remaining target datasets. This evaluation follows the label-efficient target-fine-tuning setting: 0.5% target pixels are used for fine-tuning, another disjoint 0.5% for validation, and the remaining pixels for testing. PCA is fitted only on source-domain training pixels and then applied directly to the target domain. This design avoids using any target-domain information for PCA fitting, preventing data leakage and ensuring a fair cross-dataset evaluation. Although the PCA subspace may differ across scenes, the source-domain whitened PCA still provides a decorrelated and compact spectral representation, and the network can further adapt to the target domain during fine-tuning.

4.2.2. Evaluation Metrics

To comprehensively and objectively evaluate the performance of the proposed model in hyperspectral camouflage target detection, we establish a multi-tiered evaluation framework from two perspectives: detection effectiveness and model lightweights. The detection effectiveness metrics include Precision, Probability of Detection (PD), False Alarm Rate (FAR), F1 score, Area Under the Precision-Recall Curve (PR-AUC), and Area Under the Receiver Operating Characteristic Curve (ROC-AUC). The lightweight and inference efficiency metrics comprise per-frame inference time (Time), Frames Per Second (FPS), the number of parameters (Parameters, in M), and floating-point operations (FLOPs, in M). The corresponding formulas are defined as follows. Let TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives, respectively:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

$$\text{PD} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

$$\text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (13)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{PD}}{\text{Precision} + \text{PD}} \quad (14)$$

It is worth clarifying that hyperspectral target detection differs fundamentally from land-cover classification. Classification assigns each pixel to one of several land-cover categories and is commonly evaluated by overall accuracy (OA) and Kappa. In contrast, target detection is a binary target-versus-background problem, where the target class is extremely rare. In our datasets, target pixels account for only 0.40–0.59% of all pixels. Under such extreme imbalance, OA is dominated by the background and would exceed 99.4% even for a trivial all-background predictor, while Kappa is also strongly affected by class prevalence. Therefore, OA and Kappa cannot reflect the ability to detect sparse targets. Instead, we adopt detection-oriented metrics such as PD, FAR, F1, and PR-AUC, which focus on the target class and are more suitable for severely imbalanced target detection. Specifically, PR-AUC concentrates on the dynamic trade-off between precision and recall for the positive class (camouflaged targets) and exhibits high sensitivity to the prevalence of negative samples, effectively mitigating the evaluation bias introduced by the majority class. Therefore, PR-AUC is widely recognized as the de facto standard for assessing detection performance under severe imbalance. In contrast, although ROC-AUC is theoretically insensitive to class priors, it suffers from a critical drawback in extreme imbalance scenarios: the false positive rate (FPR) is overwhelmingly dominated by the vast number of true negatives (TN). Consequently, even substantial fluctuations in the

recall of minority targets may be heavily diluted in the FPR calculation, leading ROC-AUC to yield overly optimistic estimates that fail to faithfully reflect the model's fine-grained discriminative capacity for camouflaged targets. To address this limitation, we not only plot the conventional 2D-ROC curves but also introduce 3D Relative Distance Curves (3D-RDC). By incorporating distance or confidence as the third dimension, the 3D-RDC provides a more nuanced characterization of the target-background separability boundary across multiple thresholds within the high-dimensional feature space, thereby compensating for the information loss inherent in 2D curves under unbalanced distributions.

Furthermore, the metrics Time, FPS, Parameters (M), and FLOPs (M) collectively form the basis for evaluating the model's engineering practicality. Time and FPS directly reflect the forward inference efficiency of the algorithm in hardware deployment, while Parameters and FLOPs quantitatively depict the model's storage footprint and computational overhead. Together, these four metrics rigorously validate that the proposed architecture achieves a favorable trade-off between detection accuracy and lightweight deployability.

Specifically, the inference time is defined as the end-to-end computational latency from the input HSI to the final binary detection map. This metric covers the entire inference pipeline, including spectral normalization, band-wise mean centering, optional PCA transformation, patch cropping, network forward propagation, similarity scoring, and threshold-based binarization. By contrast, training, validation, threshold tuning, and disk data loading are excluded from the time measurement. The FPS is calculated as the number of scenes processed per second, formulated as $\text{FPS} = 1/\text{Time}$. All runtime results are tested on the unified hardware and software platform specified in Section 4.2.3. All evaluations were performed with CUDA synchronization enabled, and the final results are averaged over five independent runs to ensure reliability.

4.2.3. Implementation Details and Comparison Methods

All experiments were implemented in PyTorch 1.8.0 with Python 3.8 and run on a workstation with dual NVIDIA RTX 2080Ti GPUs, each with 11 GB of memory and dual Intel Xeon E5-2698 v4 CPUs. The proposed model is optimized with AdamW, an initial learning rate of 3×10^{-4} , weight decay of 5×10^{-4} , batch size of 64, and cosine annealing over 50 epochs. Validation PR-AUC is used as the model-selection criterion, with F1 as the tie-breaker. All methods were independently run five times, and mean values, together with standard deviations where available, are reported.

To comprehensively evaluate the effectiveness of the proposed method, eight mainstream hyperspectral target detection algorithms are selected as comparison baselines. According to their technical principles, these methods can be categorized into three groups: signal processing methods, including CEM, OSP, ACE and S-ACE; sparse representation and topology-based methods, including Chessboard-shaped Topology for Hyperspectral Target Detection (CTTD); and deep learning-based methods, including UEML, TSTTD, and HTD-Mamba. These methods span the complete technical spectrum from classical statistical models to state-of-the-art deep learning frameworks, providing a sufficient comparison basis for comprehensively validating the superiority of our proposed method.

For fair comparison, the key hyperparameters of all competing methods are tuned to the optimal values recommended in their original papers. For traditional sparse detectors (CEM, OSP, ACE, S-ACE, and CTTD), the target prior spectrum is obtained by averaging 10 pure target pixels selected from the ground-truth map. For TSTTD, training samples are synthesized by mixing the prior target spectrum with all background spectra at mixing ratios ranging from 0 to 10%. Its batch size is set to 64, the learning rate is 10^{-4} , and weight decay is 0. For HTD-Mamba, the batch size is 80, the learning rate is 10^{-4} , weight decay is 10^{-4} , and the model is trained for 200 epochs with a temperature parameter of 0.1.

4.2.4. Parameter Analysis

The patch size and the number of PCA channels constitute two critical hyperparameters that jointly determine the spatial–spectral receptive field and the computational footprint of the proposed model. To ensure a fair and unbiased selection, all hyperparameter analyses are conducted exclusively on the validation set, with no access to test-set information.

We investigated the influence of the spatial patch size by searching over {3, 5, 7, 9, 11, 13}. As shown in Figure 3, both validation PR-AUC and F1 exhibit a clear upward trend as the patch size increases from 3 to 9, after which performance plateaus and then gradually declines. This behavior can be interpreted as follows: a small patch (3×3 or 5×5) preserves the spectral fidelity of the central pixel but fails to supply sufficient spatial contextual cues, making it difficult to distinguish between spectrally similar targets and backgrounds. Conversely, an excessively large patch (11×11 or 13×13) inevitably incorporates more heterogeneous background materials and mixed pixels into the receptive field, thereby introducing nuisance spectral variability. Considering the trade-off between discriminative spatial context and detection accuracy, a patch size of 9 is adopted as the optimal configuration.

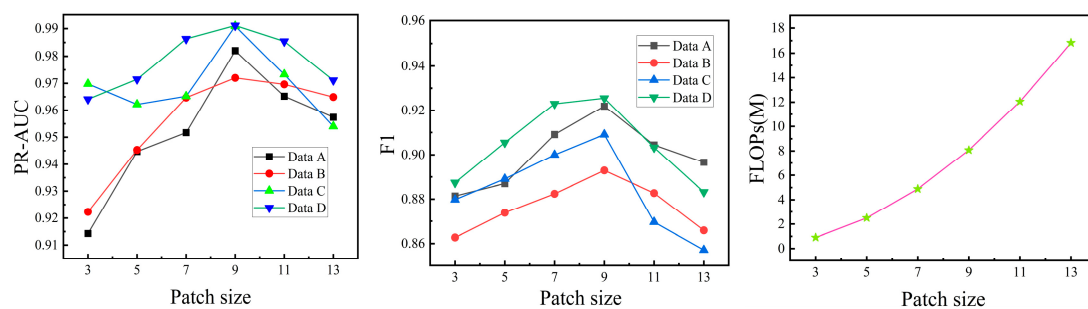


Figure 3. Variation trends of PR-AUC, F1 and FLOPs with increasing patch size.

We further evaluated the impact of the number of PCA channels over {8, 16, 24, 32, 40, 48}. As depicted in Figure 4, the detection performance improves steadily as the PCA channel count increases from 8 to 32, which is primarily attributed to the preservation of more discriminative variance that effectively separates target and background distributions. However, when the number of PCA channels exceeds 32, the performance gains tend to saturate and even exhibit a slight decline, whereas the model's parameter count and FLOPs continue to grow, incurring unnecessary computational burden without yielding corresponding accuracy improvements. Therefore, the number of PCA channels is set to 32, which strikes the most favorable balance between representational capacity and model compactness.

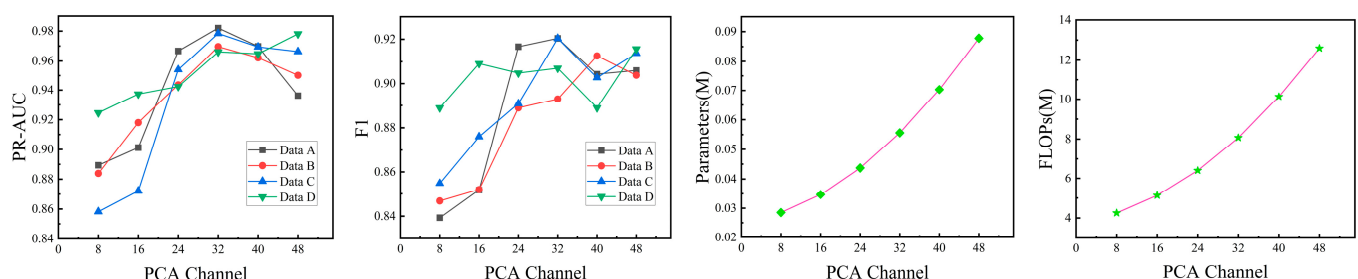


Figure 4. Variation trends of PR-AUC, F1, Parameters and FLOPs with increasing PCA channel.

In summary, based on the above systematic hyperparameter ablation, the final configuration is determined as a 9×9 spatial patch combined with 32 PCA channels, which collectively deliver the best achievable accuracy–efficiency trade-off for the target detection task.

4.3. Detection Performance Comparison

Table 2 reports quantitative results on the four datasets. The proposed model achieves the best F1, PD, PR-AUC, and ROC-AUC on all four datasets, while maintaining the lowest FAR. Its F1 scores range from 0.9438 to 0.9601, PD from 0.9465 to 0.9667, PR-AUC from 0.9814 to 0.9938, and ROC-AUC from 0.9995 to 0.9997. Meanwhile, FAR remains between 0.0002 and 0.0003. These results indicate that the proposed method does not improve recall by tolerating more false alarms; instead, it maintains a favorable precision-recall balance.

Compared with the strongest baseline in PR-AUC on each dataset, the proposed method achieves absolute improvements of 0.0330–0.2407. The improvement is particularly large on Data A, where the proposed PR-AUC reaches 0.9906. Classical detectors are unstable across scenes: some obtain high precision but low detection probability, whereas others produce many false alarms in spectrally coupled backgrounds. Deep baselines generally improve PD but suffer from low precision and high FAR under extreme class imbalance. For example, HTD-Mamba obtains PD values of 0.7879, 0.9062, 0.9334, and 0.9237 on Data A–D, but its F1 scores are only 0.2270, 0.2170, 0.4886, and 0.3075. In contrast, the proposed method jointly maintains high PD, high precision, and low FAR.

The detection maps in Figure 5 provide consistent qualitative evidence. The proposed method produces more complete target responses with fewer scattered background alarms than the compared methods. The improvement is especially visible in cluttered jungle and bush scenes, where target and background spectra overlap strongly. The 2D-ROC and 3D-RDC results in Figure 6 further show that the proposed method yields a wider separation between target and background scores across thresholds, explaining its high PR-AUC and low FAR under class imbalance.

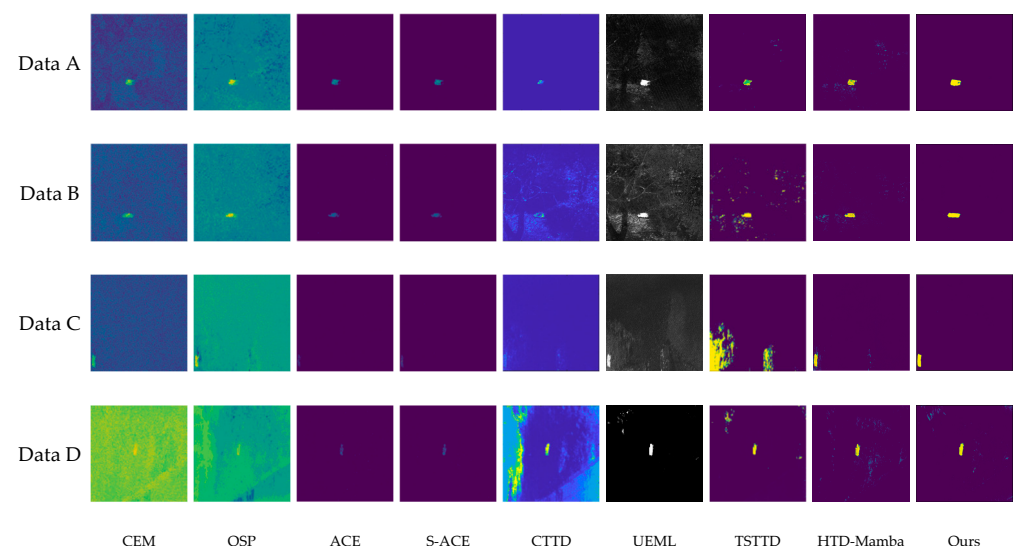


Figure 5. Visualization results of different methods on four camouflaged object datasets.

Table 2. Metrics of different methods on 4 camouflaged target datasets. The best results are shown in bold.

Datasets	Metrics	CEM	OSP	ACE	S-ACE	CTTD	UEML	TSTTD	HTD-Mamba	Ours
Data A	Precision	0.7754 ± 0.3953	0.8282 ± 0.1375	0.7406 ± 0.4117	0.7523 ± 0.4146	0.6745 ± 0.4055	0.3063 ± 0.0213	0.2944 ± 0.0218	0.1326 ± 0.0016	0.9561 ± 0.0181
	F1	0.6251 ± 0.3277	0.7360 ± 0.1471	0.3511 ± 0.3501	0.3481 ± 0.3526	0.5623 ± 0.2981	0.4352 ± 0.0344	0.4158 ± 0.0124	0.2270 ± 0.0112	0.9546 ± 0.0141
	PD	0.5236 ± 0.0774	0.6623 ± 0.0244	0.2301 ± 0.1296	0.2264 ± 0.1244	0.4821 ± 0.1395	0.7521 ± 0.0768	0.7074 ± 0.2250	0.7879 ± 0.0534	0.9541 ± 0.0341
	FAR	0.2565 ± 0.0583	0.4144 ± 0.0485	0.0081 ± 0.0020	0.0041 ± 0.0011	0.0297 ± 0.0649	0.0100 ± 0.0003	0.0100 ± 0.0057	0.0304 ± 0.0072	0.0003 ± 0.0001
	PR-AUC	0.6575 ± 0.3501	0.7499 ± 0.1644	0.6268 ± 0.3484	0.6491 ± 0.3584	0.4698 ± 0.2743	0.7280 ± 0.0708	0.6790 ± 0.0314	0.7344 ± 0.0787	0.9906 ± 0.0048
	ROC-AUC	0.9191 ± 0.1248	0.9695 ± 0.0183	0.8649 ± 0.2011	0.9123 ± 0.1381	0.8340 ± 0.2292	0.9328 ± 0.0347	0.8975 ± 0.1347	0.9134 ± 0.0256	0.9997 ± 0.0001
	Time	0.1183 ± 0.0442	0.0922 ± 0.0368	0.9539 ± 0.8558	0.7235 ± 0.0164	0.5750 ± 0.0279	9.6402 ± 0.2854	14.2761 ± 0.8457	9.0613 ± 0.7575	7.3764 ± 0.7097
	FPS	8.4531	10.8460	1.0483	1.3822	1.7391	0.1037	0.0700	0.1104	0.1356
Data B	Precision	0.9207 ± 0.0177	0.9203 ± 0.0209	0.8021 ± 0.0876	0.8491 ± 0.0367	0.5595 ± 0.1659	0.2890 ± 0.0168	0.2667 ± 0.0185	0.1233 ± 0.0024	0.9539 ± 0.0174
	F1	0.6998 ± 0.0281	0.7964 ± 0.0255	0.2777 ± 0.0772	0.3060 ± 0.0632	0.4340 ± 0.1138	0.4259 ± 0.0267	0.3900 ± 0.1007	0.2170 ± 0.0106	0.9601 ± 0.0052
	PD	0.5644 ± 0.0203	0.6956 ± 0.0293	0.1687 ± 0.0626	0.1867 ± 0.0730	0.3543 ± 0.0861	0.8099 ± 0.0611	0.7252 ± 0.0453	0.9062 ± 0.0514	0.9667 ± 0.0145
	FAR	0.2783 ± 0.0302	0.4058 ± 0.0429	0.0092 ± 0.0005	0.0046 ± 0.0003	0.0014 ± 0.0006	0.0149 ± 0.0001	0.0188 ± 0.0016	0.0323 ± 0.0052	0.0002 ± 0.0001
	PR-AUC	0.8050 ± 0.0356	0.8986 ± 0.0267	0.6802 ± 0.0887	0.7354 ± 0.0691	0.3040 ± 0.1526	0.7788 ± 0.0626	0.5576 ± 0.0384	0.8238 ± 0.0498	0.9922 ± 0.0031
	ROC-AUC	0.9771 ± 0.0052	0.9900 ± 0.0035	0.9526 ± 0.0120	0.9742 ± 0.0069	0.9330 ± 0.0222	0.9723 ± 0.0117	0.9573 ± 0.0064	0.9856 ± 0.0523	0.9997 ± 0.0001
	time	0.1282 ± 0.0553	0.1011 ± 0.0462	1.1378 ± 1.2083	1.3584 ± 1.3899	0.5197 ± 0.0118	12.4962 ± 0.2046	14.7169 ± 0.7925	8.9791 ± 0.0575	7.7413 ± 0.5786
	FPS	7.8003	9.8912	0.8789	0.7362	1.9242	0.0800	0.0679	0.1114	0.1292
Data C	Precision	0.9701 ± 0.0077	0.9827 ± 0.0082	0.9124 ± 0.0963	0.9150 ± 0.0808	0.7854 ± 0.0905	0.3070 ± 0.0097	0.3136 ± 0.0072	0.3240 ± 0.0220	0.9664 ± 0.0241
	F1	0.7417 ± 0.0283	0.8614 ± 0.0195	0.4483 ± 0.1113	0.4600 ± 0.1113	0.7264 ± 0.0602	0.4530 ± 0.0159	0.4506 ± 0.0266	0.481 ± 0.00687	0.9589 ± 0.0091
	PD	0.6005 ± 0.0244	0.7671 ± 0.0253	0.2968 ± 0.1328	0.3068 ± 0.1382	0.6753 ± 0.0532	0.8661 ± 0.0423	0.7971 ± 0.0477	0.9334 ± 0.0347	0.9521 ± 0.0128
	FAR	0.2006 ± 0.0409	0.4964 ± 0.0378	0.0070 ± 0.0015	0.0034 ± 0.0007	0.0009 ± 0.0005	0.0172 ± 0.0001	0.0076 ± 0.0008	0.0095 ± 0.0019	0.0002 ± 0.0001
	PR-AUC	0.9539 ± 0.0267	0.8989 ± 0.0208	0.8591 ± 0.1041	0.8865 ± 0.0831	0.7528 ± 0.0940	0.8841 ± 0.1372	0.9506 ± 0.0248	0.9608 ± 0.0758	0.9938 ± 0.0046
	ROC-AUC	0.9925 ± 0.0040	0.9579 ± 0.0070	0.9820 ± 0.0100	0.9895 ± 0.0061	0.9953 ± 0.0025	0.9882 ± 0.0021	0.9697 ± 0.0064	0.9896 ± 0.0048	0.9995 ± 0.0002
	time	0.8435 ± 0.1113	0.0868 ± 0.0214	0.6695 ± 0.0199	0.8640 ± 0.0197	0.5977 ± 0.0151	14.5546 ± 6.4293	17.0754 ± 1.2227	11.6554 ± 0.0228	8.3045 ± 0.6381
	FPS	1.1855	11.5207	1.4937	1.1574	1.6731	0.0687	0.0586	0.0858	0.1204
Data D	Precision	0.6260 ± 0.0272	0.9502 ± 0.0105	0.8277 ± 0.0486	0.8482 ± 0.0298	0.0726 ± 0.0042	0.4027 ± 0.0568	0.2651 ± 0.0244	0.1829 ± 0.0296	0.9423 ± 0.0236
	F1	0.6968 ± 0.0194	0.7892 ± 0.0042	0.3209 ± 0.0364	0.3221 ± 0.0362	0.1264 ± 0.0094	0.5635 ± 0.0441	0.4100 ± 0.0174	0.3075 ± 0.0187	0.9438 ± 0.0116
	PD	0.7842 ± 0.0785	0.6736 ± 0.0420	0.1989 ± 0.0176	0.1989 ± 0.0176	0.4995 ± 0.0991	0.9392 ± 0.0107	0.9043 ± 0.0403	0.9237 ± 0.0279	0.9465 ± 0.0312
	FAR	0.6843 ± 0.0876	0.5089 ± 0.0363	0.0102 ± 0.0002	0.0053 ± 0.0002	0.0253 ± 0.0032	0.0133 ± 0.0066	0.0025 ± 0.0006	0.0172 ± 0.0018	0.0002 ± 0.0001
	PR-AUC	0.5717 ± 0.0209	0.8273 ± 0.0075	0.7691 ± 0.0386	0.8175 ± 0.0363	0.0338 ± 0.0024	0.0100 ± 0.0107	0.7946 ± 0.0316	0.9073 ± 0.0304	0.9814 ± 0.0143
	ROC-AUC	0.9869 ± 0.0025	0.9898 ± 0.0006	0.9844 ± 0.0046	0.9922 ± 0.0024	0.9294 ± 0.0009	0.9518 ± 0.0124	0.9597 ± 0.0051	0.9972 ± 0.0057	0.9996 ± 0.0002
	time	0.1113 ± 0.0128	0.0851 ± 0.0181	0.6555 ± 0.0408	0.7933 ± 0.0325	0.4639 ± 0.0228	11.3640 ± 0.0942	14.4758 ± 0.0759	9.1357 ± 0.6027	7.9050 ± 0.3835
	FPS	8.9847	11.7509	1.5256	1.2606	2.1556	0.0880	0.0691	0.1095	0.1265
Parameters (M)							0.8226	0.6726	0.3247	0.0556
FLOPs (M)							12.8469	124.7253	14.0517	8.0699

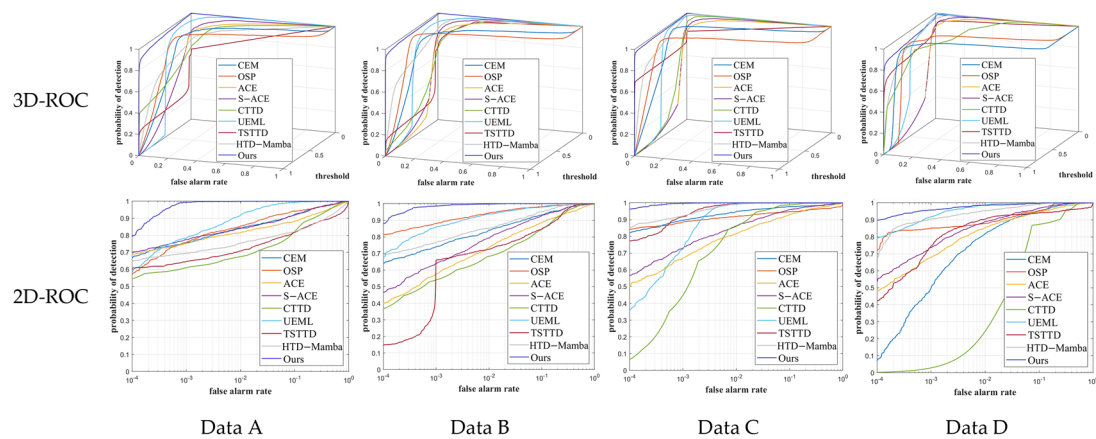


Figure 6. ROC curves of different methods on 4 camouflage target datasets.

Table 2 also reports model complexity. The proposed method contains only 0.0556 M parameters and requires 8.0699 M FLOPs per patch. Compared with UEML, it reduces parameters and FLOPs by approximately 93.2% and 37.2%, respectively. Compared with TSTTD, the reductions are 91.7% and 93.5%; compared with HTD-Mamba, they are 82.9% and 42.6%. These results show that the raw-PCA dual-branch design and depthwise multi-scale fusion retain useful spectral–spatial evidence without the parameter and computation burden of heavier deep detectors. Whole-image inference takes 7.3764–8.3045 s, corresponding to 0.1204–0.1356 scenes/s. This is faster than the deep-learning baselines listed in Table 2, although classical methods are faster because they do not perform dense patch-based deep inference.

It should be noted that the reported FLOPs are computed per patch with a patch size of 9×9 . To reflect full-image deployment, the total computational cost can be estimated as the number of pixels multiplied by the per-patch FLOPs. For a 500×500 image with 250,000 pixels, the total full-image FLOPs are approximately 2.02 T FLOPs. For Data C, the total is approximately 2.40 T FLOPs. Although the overlapping patches introduce redundant convolutions, the model contains only 0.0556 M parameters, and the whole-image inference time is 7.38–8.30 s, corresponding to a pixel-level throughput of about 30,000–34,000 pixels/s. This indicates that the proposed method is suitable for offline hyperspectral analysis and post-processing in ground reconnaissance. Future work will explore fully convolutional or strided inference to reduce patch-overlap redundancy.

4.4. Ablation Studies

Table 3 isolates the contributions of the target-aware spatial gate and the center-aware decision component. Compared with the variant without either component, the full model improves PR-AUC by 0.0366, 0.0244, 0.0205, and 0.0196 on Data A–D, respectively, and reduces FAR from 0.0009–0.0020 to 0.0002–0.0003.

The target-aware gate explains the consistent FAR reduction. In a 9×9 patch, most pixels belong to the background, and many background pixels are spectrally similar to the target. If all spatial locations are pooled with similar importance, hard-negative background evidence can leak into the global representation and produce confident false alarms. The gate estimates a spatial confidence map after raw-PCA fusion and attenuates locations that remain background-like after multi-scale contextual evidence has been integrated. Consequently, global average pooling summarizes a cleaner target-context representation rather than an unweighted mixture of target and background spectra. This mechanism is most valuable under severe class imbalance: even a small per-pixel false-alarm rate can generate many false positives because background pixels dominate the scene. The observed

FAR reduction by roughly three to ten times therefore directly supports the intended role of the gate.

Table 3. Ablation results of the gate module and center-aware component on four hyperspectral camouflaged target datasets. The best results are shown in bold.

Gate	Center	Metrics	Data A	Data B	Data C	Data D
X	X	Precision	0.9022 ± 0.0424	0.9264 ± 0.0214	0.8984 ± 0.0298	0.8856 ± 0.0147
		F1	0.9190 ± 0.0168	0.9215 ± 0.0301	0.9133 ± 0.0123	0.9110 ± 0.0164
		PD	0.9375 ± 0.0293	0.9168 ± 0.0401	0.9287 ± 0.0059	0.9378 ± 0.0338
		FAR	0.0009 ± 0.0003	0.0011 ± 0.0002	0.0015 ± 0.0004	0.0020 ± 0.0003
		PR-AUC	0.9540 ± 0.0035	0.9678 ± 0.0194	0.9733 ± 0.0022	0.9618 ± 0.0021
		ROC-AUC	0.9978 ± 0.0021	0.9985 ± 0.0010	0.9957 ± 0.0024	0.9968 ± 0.0017
		Time	6.9938 ± 0.3100	7.2768 ± 0.5455	7.8062 ± 0.4006	7.6307 ± 0.7594
✓	X	Precision	0.8937 ± 0.0217	0.9446 ± 0.0382	0.9248 ± 0.0461	0.9170 ± 0.0216
		F1	0.9215 ± 0.0016	0.9288 ± 0.0211	0.9351 ± 0.0134	0.9285 ± 0.0073
		PD	0.9510 ± 0.0266	0.9140 ± 0.0120	0.9456 ± 0.0215	0.9403 ± 0.0092
		FAR	0.0007 ± 0.0002	0.0005 ± 0.0002	0.0007 ± 0.0002	0.0008 ± 0.0001
		PR-AUC	0.9749 ± 0.0098	0.9724 ± 0.0207	0.9815 ± 0.0031	0.9854 ± 0.0049
		ROC-AUC	0.9994 ± 0.0003	0.9988 ± 0.0015	0.9991 ± 0.0003	0.9984 ± 0.0003
		Time	7.1551 ± 0.5081	7.5091 ± 0.4443	8.0554 ± 0.7319	7.6678 ± 0.6218
X	✓	Precision	0.9376 ± 0.0069	0.9013 ± 0.0388	0.9537 ± 0.0254	0.9120 ± 0.0514
		F1	0.9143 ± 0.0066	0.9113 ± 0.0313	0.9562 ± 0.0079	0.9202 ± 0.0189
		PD	0.8922 ± 0.0093	0.9216 ± 0.0257	0.9494 ± 0.0224	0.9335 ± 0.0689
		FAR	0.0007 ± 0.0002	0.0005 ± 0.0002	0.0004 ± 0.0001	0.0004 ± 0.0003
		PR-AUC	0.9572 ± 0.0126	0.9545 ± 0.0331	0.9882 ± 0.0069	0.9779 ± 0.0075
		ROC-AUC	0.9978 ± 0.0023	0.9954 ± 0.0071	0.9983 ± 0.0010	0.9986 ± 0.0007
		Time	7.0813 ± 0.6327	7.4316 ± 0.8440	7.9723 ± 0.5645	7.5888 ± 0.3960
✓	✓	Precision	0.9561 ± 0.0181	0.9539 ± 0.0174	0.9664 ± 0.0241	0.9423 ± 0.0236
		F1	0.9546 ± 0.0141	0.9601 ± 0.0052	0.9589 ± 0.0091	0.9438 ± 0.0116
		PD	0.9541 ± 0.0341	0.9667 ± 0.0145	0.9521 ± 0.0128	0.9465 ± 0.0312
		FAR	0.0003 ± 0.0001	0.0002 ± 0.0001	0.0002 ± 0.0001	0.0002 ± 0.0001
		PR-AUC	0.9906 ± 0.0048	0.9922 ± 0.0031	0.9938 ± 0.0046	0.9814 ± 0.0143
		ROC-AUC	0.9997 ± 0.0001	0.9997 ± 0.0001	0.9995 ± 0.0002	0.9996 ± 0.0002
		Time	7.3764 ± 0.7097	7.7413 ± 0.5786	8.3045 ± 0.6381	7.9050 ± 0.3835

The label of a patch refers to its central pixel, whereas global average pooling may dilute a small target among many background pixels. By explicitly concatenating the center feature with the gated global representation, the classifier receives both patch-level context and pixel-specific evidence. Adding the center component to the gate model improves F1 by 0.0289, 0.0313, 0.0149, and 0.0011 on the four datasets, indicating that center evidence is particularly helpful for preserving weak but spatially aligned target responses. Conversely, adding the gate to the center-only model improves PR-AUC by 0.0334, 0.0377, 0.0056, and 0.0035, showing that center evidence alone is insufficient when the central spectrum is ambiguous and contextual suppression is still required.

Table 3 also reports the full-image inference time. The complete model takes 7.38–8.30 s, while its variants with the gate or center component removed require 6.99–7.81 s, showing a marginal difference. As described in Section 3.2, the gate consists merely of two 1×1 convolutions followed by a sigmoid activation. As introduced in Section 3.4, the center-aware head only extracts and concatenates central features before the classifier. Therefore, the additional computational overhead is negligible. Furthermore, the full model achieves the best results across all metrics on four datasets, except for Dataset D, where its PR-AUC is slightly lower than that of the gate-only variant. This phenomenon

arises from the large spectral distance between targets and backgrounds in Dataset D; under this scenario, the standalone gate is sufficient to yield strong background suppression.

Table 4 evaluates the raw-spectral and PCA-based branches. Neither branch alone is consistently optimal. The raw branch attains PR-AUC values of 0.9288, 0.9231, 0.9862, and 0.9770, whereas the PCA branch reaches 0.9674, 0.9591, 0.9824, and 0.9727. The PCA branch is superior on Data A and Data B, while the raw branch is superior on Data C and Data D. This pattern is consistent with the spectral characteristics of the datasets. Data A and Data B have smaller target–background SADs of 0.218 and 0.053 rad, meaning that useful target evidence is weak and highly entangled with background variation in the original band space. PCA with whitening decorrelates the highly correlated bands and compresses the spectral vector into a lower-dimensional representation, which reduces the estimation burden when only 1% pixels are labeled.

Table 4. Ablation experimental results of the raw-spectral branch and PCA-based feature branch on four datasets. The best results are shown in bold.

Raw Branch	PCA Branch	Metrics	Data A	Data B	Data C	Data D
✓	X	Precision	0.8983 ± 0.0350	0.8769 ± 0.0205	0.9412 ± 0.0459	0.9027 ± 0.0458
		F1	0.8790 ± 0.0314	0.8818 ± 0.0331	0.9446 ± 0.0147	0.9226 ± 0.0067
		PD	0.8642 ± 0.0744	0.8867 ± 0.0857	0.9496 ± 0.0267	0.9459 ± 0.0342
		FAR	0.0006 ± 0.0002	0.0006 ± 0.0002	0.0005 ± 0.0002	0.0007 ± 0.0003
		PR-AUC	0.9288 ± 0.0333	0.9231 ± 0.0590	0.9862 ± 0.0106	0.9770 ± 0.0092
		ROC-AUC	0.9886 ± 0.0188	0.9881 ± 0.0193	0.9975 ± 0.0005	0.9989 ± 0.0004
		Time	4.7783 ± 0.4406	4.9700 ± 0.6585	5.2166 ± 0.6102	5.0849 ± 0.5676
X	✓	Precision	0.9360 ± 0.0413	0.9232 ± 0.0528	0.9135 ± 0.0263	0.9177 ± 0.0110
		F1	0.9112 ± 0.0425	0.9109 ± 0.0141	0.9274 ± 0.0162	0.9277 ± 0.0028
		PD	0.8877 ± 0.0445	0.9011 ± 0.0254	0.9417 ± 0.0075	0.9380 ± 0.0118
		FAR	0.0009 ± 0.0002	0.0007 ± 0.0003	0.0011 ± 0.0003	0.0005 ± 0.0002
		PR-AUC	0.9674 ± 0.0254	0.9591 ± 0.0197	0.9824 ± 0.0103	0.9727 ± 0.0039
		ROC-AUC	0.9985 ± 0.0008	0.9977 ± 0.0018	0.9987 ± 0.0011	0.9977 ± 0.0016
		Time	4.3931 ± 0.6972	4.1561 ± 0.3018	4.8201 ± 0.5417	4.6363 ± 0.3209
✓	✓	Precision	0.9561 ± 0.0181	0.9539 ± 0.0174	0.9664 ± 0.0241	0.9423 ± 0.0236
		F1	0.9546 ± 0.0141	0.9601 ± 0.0052	0.9589 ± 0.0091	0.9438 ± 0.0116
		PD	0.9541 ± 0.0341	0.9667 ± 0.0145	0.9521 ± 0.0128	0.9465 ± 0.0312
		FAR	0.0003 ± 0.0001	0.0002 ± 0.0001	0.0002 ± 0.0001	0.0002 ± 0.0001
		PR-AUC	0.9906 ± 0.0048	0.9922 ± 0.0031	0.9938 ± 0.0046	0.9814 ± 0.0143
		ROC-AUC	0.9997 ± 0.0001	0.9997 ± 0.0001	0.9995 ± 0.0002	0.9996 ± 0.0002
		Time	7.3764 ± 0.7097	7.7413 ± 0.5786	8.3045 ± 0.6381	7.9050 ± 0.3835

The dual-branch model outperforms the better single branch by 0.0232, 0.0331, 0.0076, and 0.0044 PR-AUC on Data A–D, respectively, and reduces FAR to 0.0002–0.0003. This improvement indicates that the two representations are complementary rather than simply different parameterizations of the same information. The raw branch preserves local spectral details and band-specific evidence, whereas the PCA branch provides a compact, decorrelated global representation that is less sensitive to redundant spectral variation. By concatenating both representations before contextual fusion and decision making, the classifier can distinguish samples that appear ambiguous when only one representation is available. In particular, the fused model is less likely to confuse a target with a background pixel when the two agree in the raw space but differ in the PCA space, or vice versa. These results justify the use of a dual raw-PCA design instead of selecting only one spectral representation.

Table 4 also reports the inference time. The single-branch variant takes approximately 4.1–5.2 s, whereas the two-branch model consumes 7.4–8.3 s, representing an increase of roughly 60–80%. This is attributed to the fact that the two-branch model runs two sets of

feature extraction branches simultaneously and fuses their outputs, while the single-branch variant only executes one sequence of convolutions and context blocks. The raw branch is slightly slower than the PCA branch, since its input contains 89 spectral bands, whereas the PCA branch takes 32 whitened PCA channels as input, which reduces the computational cost of the first 1×1 convolution. Despite the rise in computational overhead, the two-branch model outperforms the single-branch counterpart across all metrics. Meanwhile, the model remains lightweight with only 0.0556 M parameters.

4.5. Few-Shot Cross-Dataset Adaptation

Table 5 reports source-domain training followed by target-domain fine-tuning with only 0.5% labeled target pixels. On the twelve off-diagonal source–target pairs, PR-AUC ranges from 0.8833 to 0.9288, PD from 0.8216 to 0.9224, and F1 from 0.8138 to 0.9251. Although performance is lower than in same-dataset detection because of background, illumination, and target-type shifts, all transferred models remain above 0.88 PR-AUC. This shows that the learned representation transfers reasonably well across terrestrial camouflage scenes with very few labeled target samples. As can be observed from Figure 7, favorable detection results are obtained for all cross-dataset camouflage detection tasks. The detected targets are accurately localized, and their contours are highly consistent with the ground-truth. Only a small number of false-alarm pixels appear in the background regions.

Table 5. Metrics for cross-dataset camouflage object detection.

	Metrics	Data A	Data B	Data C	Data D
Data A	PR-AUC		0.9057	0.9140	0.9210
	PD	—	0.8577	0.9053	0.8645
	FAR		0.0023	0.0015	0.0029
	F1		0.8977	0.9251	0.8517
Data B	PR-AUC	0.9169		0.9161	0.9239
	PD	0.9224	—	0.8677	0.9151
	FAR	0.0045		0.0032	0.0039
	F1	0.9180		0.8686	0.8790
Data C	PR-AUC	0.9100	0.9152		0.9278
	PD	0.8216	0.8853	—	0.9193
	FAR	0.0033	0.0037		0.0040
	F1	0.8618	0.8758		0.8515
Data D	PR-AUC	0.8892	0.8833	0.9288	
	PD	0.8694	0.8568	0.8988	—
	FAR	0.0003	0.0073	0.0008	
	F1	0.8478	0.8138	0.8648	

The strongest transfer is from D to C, which achieves the highest PR-AUC of 0.9288, while the transfer from A to C yields the maximum F1 of 0.9251. The weakest transfer is from D to B, with a PR-AUC of 0.8833, an F1 of 0.8138, and a FAR of 0.0073. This degradation can be attributed to the substantial domain shift between Data D and Data B, which differ markedly in both background type and target spectral signatures. Overall, these results demonstrate that the proposed model generalizes across varied spectral backgrounds, while also revealing that non-negligible performance degradation persists when the source and target distributions are considerably divergent.

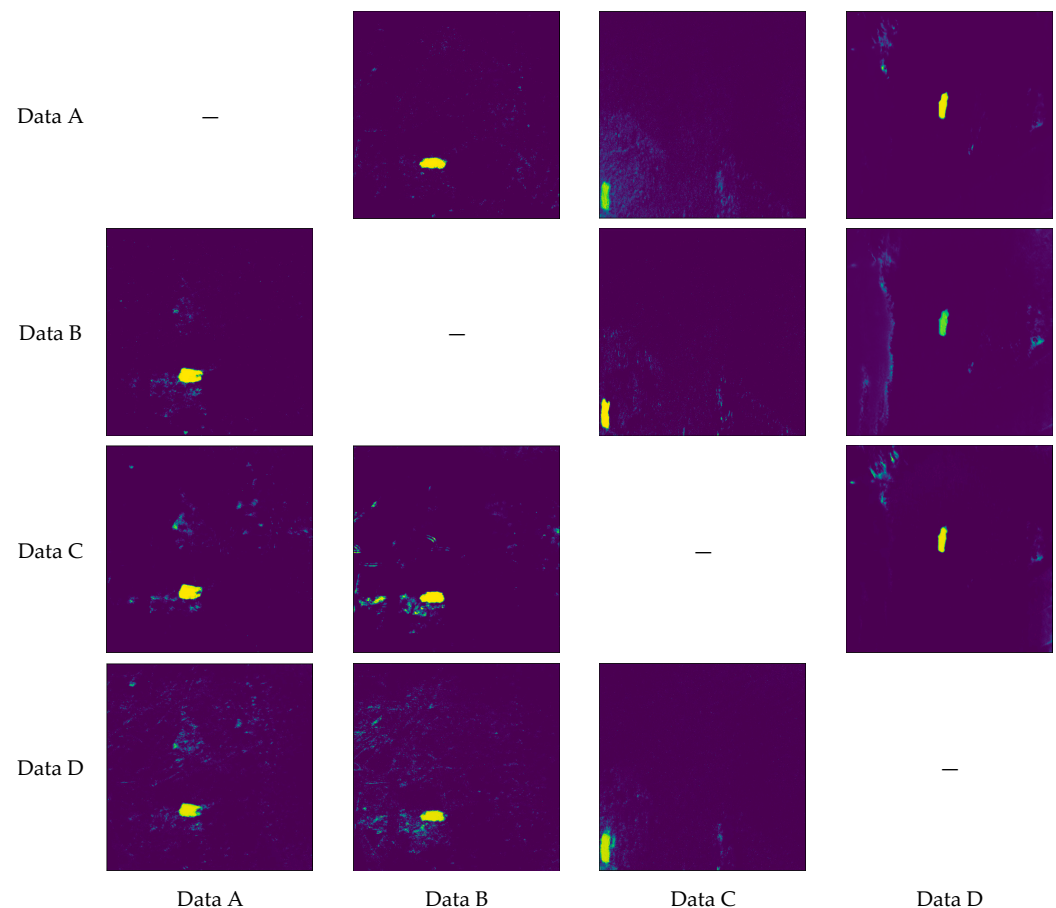


Figure 7. Visualization of cross-dataset camouflage object detection.

5. Conclusions

This paper presented a lightweight dual-branch target-aware network for land-based hyperspectral camouflage target detection. The proposed model combines a raw-spectral branch with a PCA-based branch to preserve complementary spectral evidence, introduces a target-aware spatial gate to suppress background-like responses, and uses a center-aware decision head to reduce the dilution of small targets during patch pooling. Class-weighted loss and balanced sampling further address the severe target–background imbalance in land-based camouflage scenes.

Experiments on four land-based hyperspectral datasets showed that the proposed method achieves F1 scores of 0.9438 to 0.9601, PR-AUC values of 0.9814 to 0.9938, and false-alarm rates of 0.0002 to 0.0003 under the 1% training, 4% validation, and 95% testing protocol, outperforming eight representative detectors. Ablation studies confirmed the complementarity of the raw and PCA branches and the effectiveness of the target-aware gate and center-aware head. In cross-dataset evaluation, fine-tuning with only 0.5% target pixels and validating with another 0.5% yielded PR-AUC values of 0.8833 to 0.9288, demonstrating robustness to scene changes with limited target-domain supervision. The detector requires only 0.0556 M parameters and 8.0699 M FLOPs per patch, indicating a favorable trade-off between detection performance and computational efficiency.

The current study is still bounded by the four scenes acquired by the HSI-300 spectrometer. Performance also degrades when the source and target domains differ considerably in background and target material, and dense full-image inference remains too slow for real-time deployment. Future work will therefore focus on accelerating full-image inference, exploring source-free or unlabeled domain adaptation, and validating the detector on larger multi-sensor and multi-season camouflage benchmarks.

Author Contributions: Conceptualization, X.S., B.Z. and J.Z.; software, X.S. and L.D.; validation, J.Y. and Q.C.; formal analysis, X.S.; investigation, X.S. and J.Z.; resources, B.Z. and J.Y.; data curation, X.S. and Q.C.; writing—original draft preparation, X.S.; writing—review and editing, B.Z., J.Y. and L.D.; supervision, B.Z. and J.Y.; project administration, B.Z.; funding acquisition, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study were acquired by the authors through field experiments and belong to the Department of Electronic and Optical Engineering, Army University of Engineering. The data are not publicly available due to institutional copyright and confidentiality restrictions. Researchers may request access to the data from the corresponding author upon reasonable request and with institutional approval. Code is available at <https://github.com/songxingxin1215/dual-branch-target-aware-network> (accessed on 10 October 2026).

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

Abbreviations

The following abbreviations are used in this manuscript:

HSI	Hyperspectral Imaging
CEM	Constrained Energy Minimization
OSP	Orthogonal Subspace Projection
ACE	Adaptive Cosine/Coherence Estimation
S-ACE	Sparse Adaptive Cosine/Coherence Estimator
UEML	Unbalanced Episode Meta-Learning
TSTTD	Triplet Spectralwise Transformer-based Target Detector
PCA	Principal Component Analysis
HTD	Hyperspectral Target Detection
SAD	Spectral Angle Distance
PD	Probability of Detection
FAR	False Alarm Rate
FPS	Frames Per Second
ROC	Receiver Operating Characteristic
AUC	Area Under Curve

References

1. Song, X.; Zhou, B.; Zhao, J. HCTDNet: A Novel Near-Real-Time Framework for Detecting Camouflaged Targets in Land-Based Hyperspectral Imagery. *Photonics* **2026**, *13*, 785. [CrossRef]
2. Hangloo, S.; Raina, S. Hyperspectral Target Detection: Recent Developments and Future Directions. *Arch. Comput. Methods Eng.* **2026**, *33*, 12153–12182. [CrossRef]
3. Niu, A.; Yan, X.; Wang, Q. DCH-Net: A Hyperspectral Object Detection Network with Differential Convolution and Spectral Gradient Fusion. *Pattern Recognit.* **2026**, *180*, 113956. [CrossRef]
4. Zhao, C.; Wang, M.; Feng, S. A Sparse and Spectral Smooth Regularized Low-Rank Tensor Decomposition Method for Hyperspectral Target Detection. *Int. J. Remote Sens.* **2022**, *43*, 4608–4629. [CrossRef]
5. Cheng, B.; Gao, Y. Research Progress on Anomaly Target Detection Algorithms for Hyperspectral Imagery. *Geocarto Int.* **2025**, *40*, 2589446. [CrossRef]
6. Gerster, R.; Stütz, P. Towards Robust Hyperspectral Target Detection via Test-Time Spectrum Adaptation. *Remote Sens.* **2025**, *17*, 2756. [CrossRef]
7. Zheng, X.; Kuang, Y.; Huo, Y. HTMNet: Hybrid Transformer-Mamba Network for Hyperspectral Target Detection. *Remote Sens.* **2025**, *17*, 3015. [CrossRef]
8. Zhao, J.; Fang, D.; Deng, J. Exploring a Hybrid Convolutional Framework for Camouflage Target Classification in Land-Based Hyperspectral Images. *CAAI Trans. Intell. Technol.* **2025**, *10*, 1559–1572. [CrossRef]
9. Wang, H.; Li, W.; Xia, X.G. Causal Hyperprompter: A Framework for Unbiased Hyperspectral Camouflaged Object Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *48*, 4942–4960. [CrossRef] [PubMed]

10. Zhao, J.; Fang, D.; Ying, J. A Camouflage Target Classification Method Based on Spectral Difference Enhancement and Pixel-Pair Features in Land-Based Hyperspectral Images. *Eng. Appl. Artif. Intell.* **2025**, *156*, 111141. [[CrossRef](#)]
11. Zhang, S.; Cao, Y.; Bai, L. Research on Camouflage Target Classification and Recognition Based on Mid-Wave Infrared Hyperspectral Imaging. *Remote Sens.* **2025**, *17*, 1475. [[CrossRef](#)]
12. Harsanyi, J.C. Detection and Classification of Subpixel Spectral Signatures in Hyperspectral Image Sequences. Master's Thesis, University of Maryland Baltimore County, Baltimore, MD, USA, 1993.
13. Harsanyi, J.C.; Chein, I.-C. Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 779–785. [[CrossRef](#)]
14. Kraut, S.; Scharf, L.L. The CFAR Adaptive Subspace Detector Is a Scale-Invariant GLRT. *IEEE Trans. Signal Process.* **1999**, *47*, 2538–2541. [[CrossRef](#)]
15. Zhang, L.; Du, B.; Zhong, Y. Hybrid Detectors Based on Selective Endmembers. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2633–2646. [[CrossRef](#)]
16. Sun, X.; Zhuang, L.; Gao, L.; Gao, H.; Sun, X.; Zhang, B. Information Retrieval with Chessboard-Shaped Topology for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5514515. [[CrossRef](#)]
17. Liu, Q.; Xu, Y.; Wu, Z. Unbalanced Episode Meta-Learning with Bi-Sparse Contrastive Network for Hyperspectral Target Detection. *Pattern Recognit.* **2026**, *170*, 112030. [[CrossRef](#)]
18. Jiao, J.; Gong, Z.; Zhong, P. Triplet Spectralwise Transformer Network for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5519817. [[CrossRef](#)]
19. Shen, D.; Zhu, X.; Tian, J.; Liu, J.; Du, Z.; Wang, H.; Ma, X. HTD-Mamba: Efficient Hyperspectral Target Detection with Pyramid State Space Model. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5507315. [[CrossRef](#)]
20. Wu, J.; Liu, R.; Wang, N. Contrastive Learning-Based Hyperspectral Image Target Detection Using a Gated Dual-Path Network. *Remote Sens.* **2025**, *17*, 2345. [[CrossRef](#)]
21. Zhang, Z.; Li, T.; Tang, X.; Hu, X.; Peng, Y. CAEVT: Convolutional Autoencoder Meets Lightweight Vision Transformer for Hyperspectral Image Classification. *Sensors* **2022**, *22*, 3902. [[CrossRef](#)] [[PubMed](#)]
22. Amilton, M.L.; Felipe, V.; Oriel, S.L. A Real-Time SVM-Based Hardware Accelerator for Hyperspectral Images Classification in FPGA. *Microprocess. Microsyst.* **2024**, *104*, 104998. [[CrossRef](#)]
23. Zhao, J.; Zhou, B.; Wang, G.; Liu, J.; Ying, J. Camouflage Target Recognition Based on Dimension Reduction Analysis of Hyperspectral Image Regions. *Photonics* **2022**, *9*, 640. [[CrossRef](#)]
24. Alkhatib, Q.M. ConvViTMamba: Efficient Multiscale Convolution, Transformer, and Mamba-Based Sequence Modelling for Hyperspectral Image Classification. *Int. J. Remote Sens.* **2026**, *47*, 5154–5193. [[CrossRef](#)]
25. Yang, C.; Duan, J.; Liu, X. A Center-Focused Transformer for Hyperspectral Image Classification. *Eng. Appl. Artif. Intell.* **2026**, *167*, 113813. [[CrossRef](#)]
26. Li, W.; Yuan, F.; Zhang, H. Hyperspectral Object Detection Based on Spatial-Spectral Fusion and Visual Mamba. *Remote Sens.* **2024**, *16*, 4482. [[CrossRef](#)]
27. Yun, T.; Yang, J.; Gao, F. Lightweight CNN-CEM for Efficient Hyperspectral Target Detection on Resource-Constrained Edge Devices. *Appl. Sci.* **2026**, *16*, 1719. [[CrossRef](#)]
28. Shahram, H.S.; Safa, K. A New Method to Detect Targets in Hyperspectral Images Based on Principal Component Analysis. *Geocarto Int.* **2022**, *37*, 2679–2697. [[CrossRef](#)]
29. Zhang, Y.; Wu, K.; Du, B.; Hu, X. Multitask Learning-Based Reliability Analysis for Hyperspectral Target Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2135–2147. [[CrossRef](#)]
30. Chen, J.; Chang, C.-I. Background-Annihilated Target-Constrained Interference-Minimized Filter (TCIMF) for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5540224. [[CrossRef](#)]
31. Wang, L.; Ji, L.; Geng, X.; Zhang, L. Comments on “Hierarchical Suppression Method for Hyperspectral Target Detection”. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6014304. [[CrossRef](#)]
32. Lin, S.; Cai, E.; Zeng, Y. A Hyperspectral Target Detection Method by Combining Robust Dictionaries and Constrained Energy Minimization Regularized Low-Rank and Sparse Representation Model. *Adv. Space Res.* **2026**, *78*, 3523–3544. [[CrossRef](#)]
33. Zhao, X.; Zhang, M.; Li, W.; Gao, K.; Tao, R. A Sparse Representation and Cauchy Distance Combination Graph for Hyperspectral Target Detection. *Remote Sens. Lett.* **2023**, *14*, 1218–1226. [[CrossRef](#)]
34. Zhou, T.; Cen, Y.; He, J. K-Means Adaptive 2DSSA Based on Sparse Representation Model for Hyperspectral Target Detection. *Infrared Phys. Technol.* **2024**, *143*, 105616. [[CrossRef](#)]
35. Pang, Y.; Qu, Q.; Wang, M. Sparse-Regularized Tensor PCA for Hyperspectral Target Detection. *Infrared Phys. Technol.* **2026**, *158*, 106820. [[CrossRef](#)]
36. Chen, B.; Liu, L.; Zou, Z.; Shi, Z. Target Detection in Hyperspectral Remote Sensing Image: Current Status and Challenges. *Remote Sens.* **2023**, *15*, 3223. [[CrossRef](#)]

37. Zhao, J.; Wang, G.; Zhou, B. Exploring an Application-Oriented Land-Based Hyperspectral Target Detection Framework Based on 3D-2D CNN and Transfer Learning. *EURASIP J. Adv. Signal Process.* **2024**, *2024*, 37. [\[CrossRef\]](#)
38. Yang, X.; Zhao, M.; Shi, S.; Chen, J. Deep Constrained Energy Minimization for Hyperspectral Target Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8049–8063. [\[CrossRef\]](#)
39. Xu, Y.; Sun, J.; Ren, P. Diffusion Models for Hyperspectral Image Superresolution: From Noise Prediction to Direct Image Prediction. *IEEE Geosci. Remote Sens. Mag.* 2026, *early access*. [\[CrossRef\]](#)
40. Feng, S.; Feng, R.; Liu, J.; Zhao, C.; Xiong, F.; Zhang, L. An Attention-Based Multiscale Spectral–Spatial Network for Hyperspectral Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5503705. [\[CrossRef\]](#)
41. 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. [\[CrossRef\]](#)
42. Xu, S.; Geng, S.; Xu, P.; Chen, Z.; Gao, H. Cognitive Fusion of Graph Neural Network and Convolutional Neural Network for Enhanced Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5515915. [\[CrossRef\]](#)
43. Li, J.; Zhou, S.; Liu, Y.; Hu, G.; Liu, X. Hyperspectral Image Classification Based on a Spatial–Spectral Dual-Branch Mamba Architecture. *Remote Sens.* **2026**, *18*, 2526. [\[CrossRef\]](#)
44. Liu, H.; Li, H.; Yu, B.; Luo, F.; Shao, G. HDMamba: A hierarchically guided dual-branch Mamba network for hyperspectral image classification. *J. Supercomput.* **2026**, *82*, 619. [\[CrossRef\]](#)
45. Li, T.; Cao, Y.; Guo, X.; Zhang, S.; Yan, L. A Dual-Branch CNN with Depthwise Separable Fusion for Hyperspectral Image Classification. *Remote Sens.* **2026**, *18*, 1685. [\[CrossRef\]](#)
46. Liu, H.; Bi, W.; Mughees, N. Enhanced Hyperspectral Image Classification Technique Using PCA-2D-CNN Algorithm and Null Spectrum Hyperpixel Features. *Sensors* **2025**, *25*, 5790. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Uddin, M.P.; Mamun, M.A.; Hossain, M.A. Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification. *Int. J. Remote Sens.* **2019**, *40*, 7190–7220. [\[CrossRef\]](#)
48. Gao, S.; Jiang, X.; Zhang, Y.; Liu, X.; Xiong, Q.; Cai, Z. Superpixelwise PCA based data augmentation for hyperspectral image classification. *Multimed. Tools Appl.* **2024**, *83*, 81209–81229. [\[CrossRef\]](#)
49. Zhan, Y.; Wang, Y.; Yu, X. Semisupervised Hyperspectral Image Classification Based on Generative Adversarial Networks and Spectral Angle Distance. *Sci. Rep.* **2023**, *13*, 22019. [\[CrossRef\]](#) [\[PubMed\]](#)

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.