

HyperCOD: The First Challenging Benchmark and Baseline for Hyperspectral Camouflaged Object Detection

Shuyan Bai, Tingfa Xu*, Peifu Liu, Yuhao Qiu, Huiyan Bai, Huan Chen, Yanyan Peng, Jianan Li*

Beijing Institute of Technology

Abstract

RGB-based camouflaged object detection struggles in real-world scenarios where color and texture cues are ambiguous. While hyperspectral image offers a powerful alternative by capturing fine-grained spectral signatures, progress in hyperspectral camouflaged object detection (HCOD) has been critically hampered by the absence of a dedicated, large-scale benchmark. To spur innovation, we introduce HyperCOD, the first challenging benchmark for HCOD. Comprising 350 high-resolution hyperspectral images, it features complex real-world scenarios with minimal objects, intricate shapes, severe occlusions, and dynamic lighting to challenge current models. The advent of foundation models like the Segment Anything Model (SAM) presents a compelling opportunity. To adapt the Segment Anything Model (SAM) for HCOD, we propose HyperSpectral Camouflage-aware SAM (HSC-SAM). HSC-SAM ingeniously reformulates the hyperspectral image by decoupling it into a spatial map fed to SAM’s image encoder and a spectral saliency map that serves as an adaptive prompt. This translation effectively bridges the modality gap. Extensive experiments show that HSC-SAM sets a new state-of-the-art on HyperCOD and generalizes robustly to other public HSI datasets. The HyperCOD dataset and our HSC-SAM baseline provide a robust foundation to foster future research in this emerging area.

Code — <https://github.com/Baishuyanyan/HyperCOD>

Datasets — <https://github.com/Baishuyanyan/HyperCOD>

Introduction

Camouflaged object detection (COD) identifies objects blending with their surroundings (2020a; 2022), with applications in medical diagnosis (2023), ecological monitoring (2019) and search-and-rescue (2020b). Existing RGB-based methods rely on spatial cues like texture and edges, but fail when camouflage is perfected or illumination flattens spatial distinctions (Fig. 1(a)(b)), exposing limitations of color-based vision.

Hyperspectral image (HSI) captures hundreds of spectral bands, revealing material differences invisible to RGB cameras (2025) (Fig. 1(c)). The unique spectral properties of

* Correspondence to: Tingfa Xu and Jianan Li.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

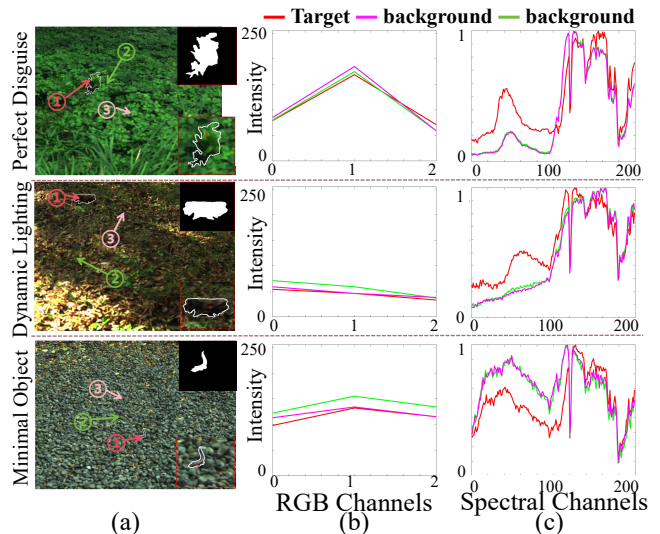


Figure 1: While RGB images show near-indistinguishable camouflage, hyperspectral signatures reveal significant differences, highlighting spectral advantages for COD.

HSI make it exceptionally well-suited for detecting camouflaged objects, especially those that are small, partially occluded, or set against cluttered backdrops. We formalize this capability as **Hyperspectral Camouflaged Object Detection (HCOD)**, a novel task currently constrained by the lack of large-scale, diverse benchmark datasets.

To catalyze research and provide a standardized evaluation platform for HCOD, we present **HyperCOD**, the first comprehensive benchmark for hyperspectral camouflaged object detection. HyperCOD features 350 high-quality HSI cubes, each with high spatial resolution (1240×1680) and dense spectral information (200 bands from 400-1000 nm). Captured across 11 diverse natural environments (grasslands, snowfields, etc.), the dataset incorporates challenging scenarios including small targets, complex occlusions, cluttered backgrounds, and dynamic lighting. It is designed not only to benchmark existing methods but also to drive the development of novel algorithms that can fully exploit spectral information for robust real-world camouflage detection.

While hyperspectral image provides rich information, ef-

fectively harnessing it remains a challenge. Traditional hyperspectral analysis methods often rely on pixel-wise, hand-crafted spectral metrics (2023), which lack spatial context and exhibit poor robustness in complex scenes (2024). On the other hand, the remarkable success of large vision models like the Segment Anything Model (SAM) in RGB domains is inspiring. However, their direct application to hyperspectral images is infeasible due to the prohibitive computational cost and the modality gap between RGB and high-dimensional hyperspectral images.

To bridge this gap, we propose **HSC-SAM**, a novel framework designed to adapt the powerful segmentation capabilities of SAM for the HCOD task. To resolve the modality mismatch, our Spectral-Spatial Decomposition Module (SSDM) decouples HSI into two complementary representations: a CIEXYZ color-space mapped image that preserves spatial structures and a spectral saliency map that highlights the camouflaged target. This design allows SAM’s RGB-pretrained encoder to process familiar spatial structures while using the spectral map as a high-quality, spectrum-driven prompt to guide segmentation.

To improve computational efficiency, our Spectral-Guided Token Dropout (SGTD) mechanism leverages the generated saliency map to selectively discard image tokens from non-salient background regions prior to the expensive attention computations. This principled filtering reduces the processing load and focuses the model’s capacity on promising foreground areas. Besides, the Fusion Detail Enhancer (FDE) re-injects high-frequency spatial features to refine object contours and ensure precise boundary localization.

Experiments on HyperCOD validate HSC-SAM’s state-of-the-art performance (MAE: 0.0017) and its effectiveness on hyperspectral salient object detection. Our HyperCOD dataset and the HSC-SAM framework provide a solid foundation for future advancements in hyperspectral camouflaged object detection. The core contributions of this work are summarized as follows:

- We introduce HyperCOD, the first large-scale and challenging dataset for hyperspectral camouflaged object detection, covering diverse scenes and multiple camouflage scenarios.
- We propose HSC-SAM, a unified spectral-aware detection framework that integrates spectral-spatial decomposition and complementary prompt learning to adapt SAM to the hyperspectral domain.
- We design Spectral-Guided Token Dropout (SGTD) and Fusion Detail Enhancer (FDE) modules to improve saliency-guided pruning and boundary refinement, resulting in precise segmentation performance.

Related Work

Camouflaged Object Detection Datasets. RGB-based camouflaged object detection has advanced with datasets like CHAMELEON (2018), CAMO-V1 (2019), COD10K (2022), and NC4K (2021). However, their limited spectral resolution struggles with foreground-background ambiguity. For hyperspectral analysis, HS-SOD (2018) was the first dataset tailored for HSOD, though limited to 60

Dataset	Modality	Resolution	Channels
CHAMELEON (2018)	RGB	450×300~2304×3456	3
CAMO-V1 (2019)	RGB	154×156~7360×4912	3
CAMO++ (2022)	RGB	Unreleased	3
COD10K (2022)	RGB	300×199~2976×3968	3
NC4K (2021)	RGB	354×268~1280×960	3
ACOD-12K (2024)	RGB-D	Unreleased	4
HyperCOD (Ours)	HSI	1680×1240	200

Table 1: Comparison of Existing COD Datasets. **HyperCOD** is the first dataset to introduce HSIs for COD.

images. HSOD-BIT (2024) and its recent version HSOD-BIT-V2 (2025) expanded the scale to 319 annotated images and introduced various challenges such as non-uniform illumination and overexposure. As these focus on saliency detection, we introduce **HyperCOD**, the first large-scale hyperspectral COD dataset featuring diverse real-world camouflage challenges, establishing a robust benchmark for advancing HCOD research.

Camouflaged Object Detection Methods. Traditional COD methods (2003; 1998; 2020) rely on hand-crafted features (*e.g.*, texture, color, motion), which struggle with complex backgrounds and variable object appearances, leading to poor robustness. Recent models (SINet (2022), D2CNet (2022)) employ attention mechanisms, while edge-guided approaches (SegMaR (2022), BGNet (2022)) integrate boundary cues. Other methods, such as ZoomNet (2022) and FSEL (2024), address modality-specific challenges via confidence-aware learning and feature fusion. Despite their success, these methods remain confined to RGB modality. Our framework overcomes this by fully leveraging hyperspectral information.

SAM-based Methods. While SAM adaptations like SAM-COD (2024a) and SAM-Adapter (2024b) have advanced COD through efficient tuning strategies, and approaches such as ProMAC (2024) and MD-SAM (2024) have explored multi-scale processing, these methods remain constrained to RGB inputs and manual prompting. Our framework addresses both limitations by automatically extracting discriminative spectral features while eliminating the need for handcrafted prompts.

HyperCOD Dataset

Our HyperCOD dataset comprises 350 high-quality hyperspectral images, each containing 200 spectral bands spanning 400–1000 nm with a spatial resolution of 1240 × 1680 pixels. The dataset is partitioned into a training set (280 samples) and a testing set (70 samples) at a 4:1 ratio.

Comparison with Existing Datasets. As shown in Tab. 1, HyperCOD surpasses existing camouflaged object detection and hyperspectral object detection datasets in several key aspects. While RGB-based COD datasets are well-established, their single-modality nature restricts the utilization of spectral information. HyperCOD is the first dataset tailored for

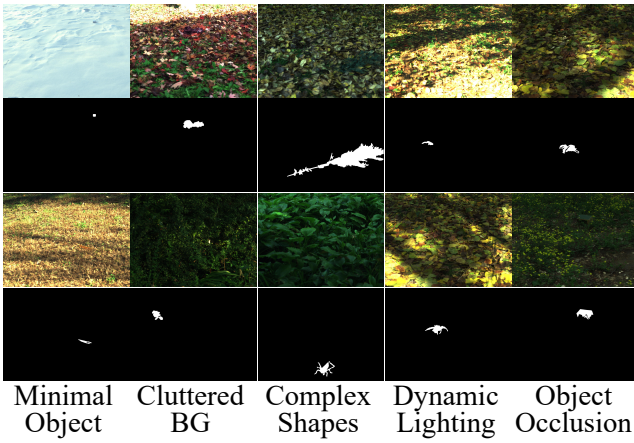


Figure 2: Representative samples and corresponding ground truth masks from five challenging scenarios.

hyperspectral COD, offering high-resolution spatial details and dense spectral coverage. It ensures greater data diversity and introduces more complex and challenging scenarios for the evaluation of robust hyperspectral camouflaged object detection models.

Data Collection. HyperCOD was constructed using the HySpex Mjolinir-1024 hyperspectral camera coupled with a STANDA 8 MR 190 precision motorized rotation stage for line-by-line scanning of natural scenes. To ensure diversity, data were captured under varying weather conditions (e.g., cloudy, sunny) and times of day (morning, afternoon), while maintaining consistent imaging parameters such as exposure time, white balance, and gain compensation.

Annotation Process. Pixel-level ground truth annotations were created on false-color images using MATLAB’s ImageLabeler toolbox. Camouflaged objects were labeled as foreground (1), while the remaining regions were marked as background (0). Each sample in the dataset includes a hyperspectral image (.mat), a binary segmentation mask (.png), and a pseudo-RGB image (.jpg).

Challenge Attributes. HyperCOD introduces significant challenges categorized into five key attributes: (i) Minimal Objects (MO) featuring extremely small and inconspicuous targets, (ii) Complex Shapes (CS) with irregular or fragmented boundaries, (iii) Dynamic Lighting (DL) under strong illumination variations, (iv) Object Occlusion (OO) with partially obscured objects, and (v) Cluttered Backgrounds (CB) in environmentally complex scenes—as depicted in Fig. 2. These attributes collectively highlight the critical role of spectral information in camouflage detection, imposing higher demands on detection models while providing a rigorous benchmark for hyperspectral frameworks.

Statistical Analysis. We perform a comprehensive statistical analysis of HyperCOD across multiple dimensions to ensure diversity, balance, and challenge.

Scene diversity. is ensured by including a wide range of natural environments. The training and testing splits are care-

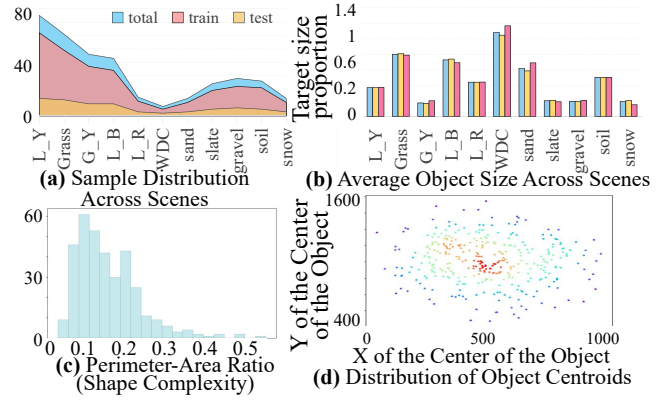


Figure 3: Statistics of the HyperCOD Dataset.

fully balanced across scene categories, as shown in Fig. 3(a).

Object size distribution. is analyzed in Fig. 3(b). Most targets in HyperCOD are relatively small, increasing the dataset’s difficulty. Samples with object-to-image area ratios below 0.5% are classified as *tiny-object cases*, with proportional representation in both training and testing subsets to maintain consistent difficulty.

Boundary complexity. is evaluated using the edge-to-perimeter ratio (Fig. 3(c)). Targets with ratios exceeding 0.3 are labeled as *complex-edge cases*, requiring robust segmentation and fine-grained contour learning. These examples encourage models to better capture subtle structural details.

Spatial distribution. is examined in Fig. 3(d), which shows the centroid distribution of target objects. HyperCOD exhibits minimal center bias, ensuring a uniform spatial distribution. This property mitigates common pitfalls in existing datasets and promotes fairer evaluation, encouraging models to generalize across diverse object positions.

Method

The goal of the hyperspectral camouflaged object detection task is to localize camouflaged objects from a hyperspectral image cube $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$, where H , W , and C denote the spatial height, width, and spectral channels, respectively, and to generate a corresponding saliency map $\mathbf{F} \in \mathbb{R}^{H \times W \times 1}$ that highlights object regions. To address this task, we propose the HyperSpectral Camouflage-aware SAM (HSC-SAM) framework, as illustrated in Fig. 4.

HSC-SAM employs a Spectral-Spatial Decomposition Module (SSDM) to extract complementary spatial and spectral features from hyperspectral imagery. The spatial map preserves structural details for the SAM image encoder, while the spectral saliency map highlights material differences for the prompt encoder. This spectral prompt further guides token pruning to suppress non-salient background activations. During decoding, spatial-spectral joint learning enhances object localization to produce initial segmentations. Finally, the Fusion Detail Enhancer (FDE) module supervises high-frequency boundary restoration during training, yielding more precise segmentation results.

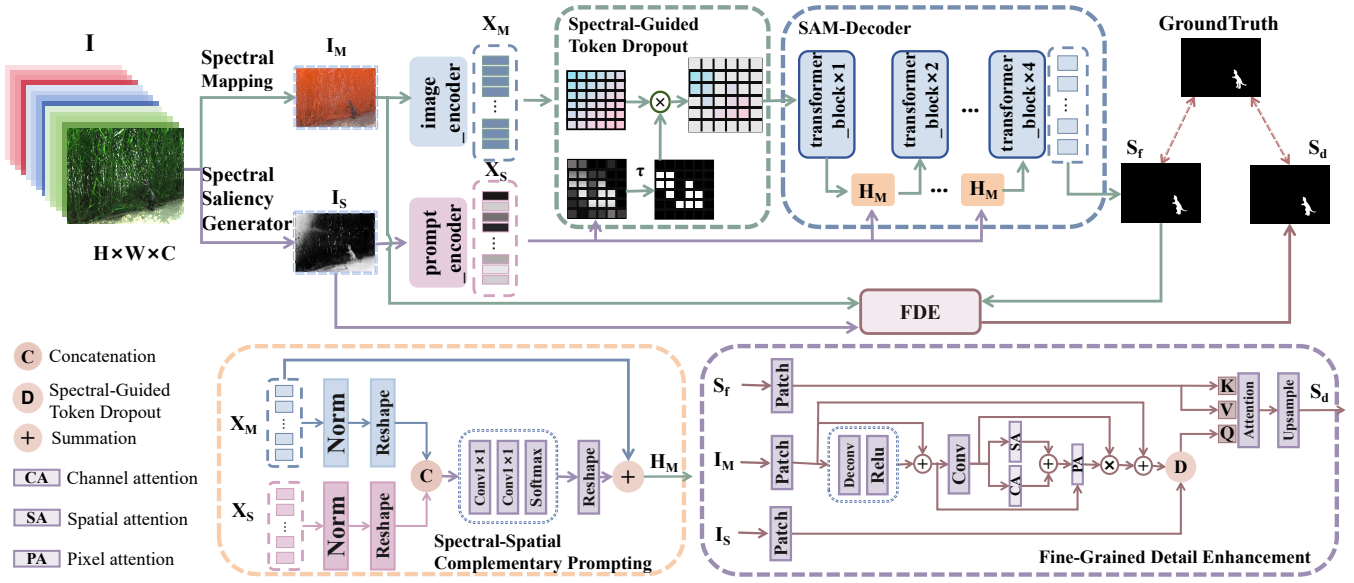


Figure 4: Overview of the proposed HSC-SAM framework. Spatial features I_M are used in the SAM image encoder, while spectral saliency I_S guides prompt encoding and token pruning. Joint spatial-spectral learning enhances object localization, and a Fusion Detail Enhancer (FDE) refines boundary details for accurate segmentation.

Spectral-Spatial Decomposition Module

Color Space Mapping. Metamerism causes spectrally distinct materials to appear identical in RGB images due to sensor limitations, posing challenges for camouflaged object detection. In contrast, the CIE XYZ color space integrates spectral reflectance with standardized color matching functions across the visible spectrum, preserving material-dependent spectral nuances.

To construct the transformed image I_M , we uniformly sample 33 representative bands from the original 200 spectral channels and apply the standard CIE color matching functions W_t (2004) to map them into the CIEXYZ color space. For a hyperspectral image $I(\lambda)$, the tristimulus values (X, Y, Z) are computed as:

$$I_M = \sum_{i=1}^N I(\lambda_i) \cdot W_t(\lambda_i), \quad t \in \{X, Y, Z\}, \quad (1)$$

where $I(\lambda_i)$ is the spectral intensity at wavelength λ_i , $W_t(\lambda_i)$ is the color matching function for channel t , and N is the number of spectral bands. This preserves material-dependent spectral signatures that RGB sensors lose.

Spectral Saliency Generator. While the aforementioned steps enhance the discriminability of foreground and background in terms of spatial structure, they do not fully exploit spectrum. To address this limitation, we employ SSG proposed by Liu *et al.* (2023) to generate a spectral prompt $I_S \in \mathbb{R}^{H \times W \times 3}$, which encodes spectral saliency cues. It constructs an N -level Gaussian pyramid from the input hyperspectral image and computes spectral angular distances (SAD) between feature vectors extracted from non-adjacent pyramid levels. For a pixel (i, j) , let $v_c^{(i,j)}$ and $v_{c+3}^{(i,j)}$ repre-

sent spectral vectors from levels c and $c + 3$, respectively. The intermediate saliency map at level c is defined as:

$$S^{(c)}(i, j) = \arccos \left(\frac{v_c^{(i,j)} \cdot v_{c+3}^{(i,j)}}{\|v_c^{(i,j)}\|_2 \cdot \|v_{c+3}^{(i,j)}\|_2} \right). \quad (2)$$

The spectral prompt I_S is then obtained by concatenating the intermediate saliency maps across levels $c \in \{2, 3, 4\}$.

In our framework, I_M is used as the encoder input, while I_S serves as a spectral prompt to guide localization.

Spectral-Guided Token Dropout

To suppress background noise and reduce computational overhead, we introduce a Spectral-Guided Token Dropout (SGTD) mechanism based on the spectral saliency map. This mechanism prunes background tokens at the token level by leveraging spectral attention, thus improving model focus on salient regions and enhancing feature discrimination.

Given the encoded image features $X_M \in \mathbb{R}^{B \times N \times C}$ and the corresponding patch-wise spectral saliency map $X_S \in \mathbb{R}^{B \times N \times C}$, we first compute the mean activation across the channel dimension to obtain a saliency score for each token:

$$S_{\text{score}} = \frac{1}{C} \sum_{c=1}^C X_S^{(:, :, c)} \in \mathbb{R}^{B \times N \times 1}, \quad (3)$$

where S_{score} denotes the scalar saliency response per token.

We then apply a user-defined threshold τ to binarize the saliency map and generate a token-wise dropout mask:

$$M_{\text{mask}}^{(i,j)} = \begin{cases} 1, & \text{if } S_{\text{score}}^{(i,j)} \geq \tau, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where i indexes the batch and j indexes the token position.

The binary mask is finally applied to the encoded image features to suppress background information:

$$\widetilde{\mathbf{X}}_M = \mathbf{X}_M \odot \mathbf{M}_{\text{mask}}, \quad (5)$$

where \odot denotes element-wise multiplication with broadcasting along the channel dimension.

This dropout operation directly integrates spectral priors into token selection, enabling the model to dynamically focus on high-saliency regions while discarding low-importance tokens, thus improving both detection precision and computational efficiency.

Spectral-Spatial Complementary Prompting

To align spectral saliency with spatial structural cues for more accurate target localization, we introduce a Spectral-Spatial Complementary Prompting mechanism. The hyperspectral branch provides rich spectral priors, enabling the model to identify subtle material variations that are imperceptible in RGB imagery, especially under challenging camouflage.

Given image tokens $\mathbf{X}_M^i \in \mathbb{R}^{B \times L \times C}$ and spectral tokens $\mathbf{X}_S \in \mathbb{R}^{B \times L \times C}$, we first apply layer normalization:

$$\widetilde{\mathbf{H}}_M^i = \text{LN}(\mathbf{X}_M^i), \quad \widetilde{\mathbf{H}}_S = \text{LN}(\mathbf{X}_S). \quad (6)$$

Each normalized token sequence is reshaped into 2D feature maps using a token-to-feature transformation $\phi: \mathbb{R}^{B \times L \times C} \rightarrow \mathbb{R}^{B \times C \times H \times W}$:

$$\mathbf{F}_{\text{img}} = \phi(\widetilde{\mathbf{H}}_M^i), \quad \mathbf{F}_{\text{spec}} = \phi(\widetilde{\mathbf{H}}_S). \quad (7)$$

The two features are concatenated along the channel dimension and fused via a prompt block $\mathcal{P}(\cdot)$, which integrates spectral saliency and spatial structure:

$$\mathbf{F}_{\text{fused}} = \mathcal{P}(\text{Concat}(\mathbf{F}_{\text{img}}, \mathbf{F}_{\text{spec}})). \quad (8)$$

The fused features are projected back to token space and added to the original image tokens in a residual manner:

$$\widehat{\mathbf{H}}_M^{i+1} = \mathbf{H}_M^i + \phi^{-1}(\mathbf{F}_{\text{fused}}). \quad (9)$$

This complementary prompting mechanism allows the model to refine spatial attention using spectral contrast, promoting fine-grained camouflage object segmentation even in low-contrast or texture-similar environments.

Fine-Grained Detail Enhancement

To recover boundary details lost in high-level semantic decoding, we introduce a Fine-Grained Detail Enhancement (FDE) module that injects low-level spatial features via joint attention modulation.

Let the spectral-mapped pseudo-RGB input be $\mathbf{I}_M \in \mathbb{R}^{B \times C \times H \times W}$. We extract structural features using depthwise separable convolution:

$$\mathbf{F} = \phi(\mathbf{I}_M), \quad \mathbf{F} \in \mathbb{R}^{B \times C \times H \times W}. \quad (10)$$

To refine \mathbf{F} , we apply three types of attention. Channel Attention (CA) computes inter-channel dependencies:

$$\mathbf{A}_{\text{CA}} = \sigma(W_2 \delta(W_1 \cdot \text{GAP}(\mathbf{F}))) \in \mathbb{R}^{B \times C \times 1 \times 1}, \quad (11)$$

Method	MAE ↓	E ↑	S ↑	Adp-F ↑	Params	FLOPs
<i>RGB-based Methods</i>						
SINet-V2	0.0033	0.732	0.746	0.480	27.0	12.2
ZoomNet	0.0044	0.831	0.757	0.352	32.4	203.5
FRINet	0.0027	0.889	0.759	0.604	100.0	112.7
SAM2-UNet	0.0022	0.899	0.805	0.641	216.4	128.4
HGINet	0.0039	0.850	0.766	0.585	400.0	530.4
Camoforner	0.0070	0.673	0.355	0.660	71.4	41.8
<i>Hyperspectral Salient Object Detection Methods</i>						
SAD	0.1505	0.325	0.483	0.0061	–	–
DMSSN	0.0295	0.687	0.446	0.409	31.0	8.1
SMN-PVT	0.0066	0.654	0.531	0.103	10.2	14.8
Hyper-HRNet	0.0149	0.818	0.608	0.150	29.6	19.0
<i>Hyperspectral Camouflage Object Detection Methods</i>						
HSC-SAM	0.0017	0.853	0.802	0.681	11.7	94.2

Table 2: Results on the HyperCOD dataset. Bold values indicate the best performance in each column.

where GAP denotes global average pooling, δ is ReLU, and σ is sigmoid, W_1 and W_2 are convolutional layers whose weights are learned during training. Spatial Attention (SA) captures boundary-localized responses:

$$\mathbf{A}_{\text{SA}} = \sigma\left(\text{Conv}_{k \times k}(\text{AvgPool}_c(\mathbf{F}) \oplus \text{MaxPool}_c(\mathbf{F}))\right), \quad (12)$$

where \oplus denotes channel-wise concatenation, $\text{Conv}_{k \times k}$ is a convolution kernel (e.g., $k = 7$), and $\mathbf{A}_{\text{SA}} \in \mathbb{R}^{B \times 1 \times H \times W}$.

$$\mathbf{A}_{\text{PA}} = \sigma(\text{Conv}_{1 \times 1}(\mathbf{F})), \quad \mathbf{A}_{\text{PA}} \in \mathbb{R}^{B \times C \times H \times W}. \quad (13)$$

The overall modulation map is computed as:

$$\mathbf{A} = \mathbf{A}_{\text{CA}} \cdot \mathbf{A}_{\text{SA}} \cdot \mathbf{A}_{\text{PA}}, \quad \mathbf{A} \in \mathbb{R}^{B \times C \times H \times W}. \quad (14)$$

We then enhance the decoder prediction $\mathbf{S}_f \in \mathbb{R}^{B \times 1 \times H \times W}$ with the modulated fine-grained features:

$$\mathbf{S}_d = \mathbf{S}_f + \mathbf{A} \odot \mathbf{F}, \quad (15)$$

where \odot denotes element-wise multiplication.

This formulation integrates boundary-aware low-level features into the segmentation pipeline, yielding sharper and more complete object delineation.

Learning Objective

We adopt a composite loss function that combines Binary Cross-Entropy (BCE) and Intersection-over-Union (IoU) to supervise both the decoder output and the final prediction. This encourages accuracy and structural consistency.

The loss for the decoder output \mathbf{S}_d is defined as:

$$\mathcal{L}_{\text{dec}} = \mathcal{L}_{\text{BCE}}(\mathbf{S}_d, \mathbf{G}) + \mathcal{L}_{\text{IoU}}(\mathbf{S}_d, \mathbf{G}), \quad (16)$$

and the loss for the final prediction \mathbf{S}_f as:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{BCE}}(\mathbf{S}_f, \mathbf{G}) + \mathcal{L}_{\text{IoU}}(\mathbf{S}_f, \mathbf{G}). \quad (17)$$

The total loss is then given by $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{dec}} + \mathcal{L}_{\text{final}}$.

Method	Cluttered Backgrounds		Complex Shapes		Dynamic Lighting		Multiple Objects		Occlusion	
	MAE ↓	Adp-F ↑	MAE ↓	Adp-F ↑	MAE ↓	Adp-F ↑	MAE ↓	Adp-F ↑	MAE ↓	Adp-F ↑
DMSSN (2024)	0.0762	0.0737	0.0809	0.0793	0.0784	0.0110	0.0818	0.0096	0.1223	0.0544
Hyper-HRNet (2025)	0.0157	0.2243	0.0161	0.2025	0.0137	0.1581	0.0151	0.0383	0.0095	0.3183
SMN (2023)	0.0105	0.0826	0.0103	0.1087	0.0068	0.0836	0.0036	0.0344	0.0081	0.2416
HGINet (2024)	0.0049	0.7451	0.0061	0.6635	0.0034	0.7277	0.0017	0.5176	0.0020	0.8486
SAM2-UNet (2024)	0.0036	0.6613	0.0029	0.6712	0.0022	0.7219	0.0010	0.5600	0.0022	0.8261
SINet-V2 (2022)	0.0054	0.5937	0.0058	0.5322	0.0036	0.4945	0.0016	0.3721	0.0028	0.7941
ZoomNet (2022)	0.0073	0.3726	0.0064	0.4511	0.0050	0.4135	0.0021	0.2028	0.0041	0.6220
HSC-SAM (Ours)	0.0033	0.7588	0.0026	0.7727	0.0015	0.8242	0.0011	0.5828	0.0012	0.8698

Table 3: Performance comparison under five challenging attributes.

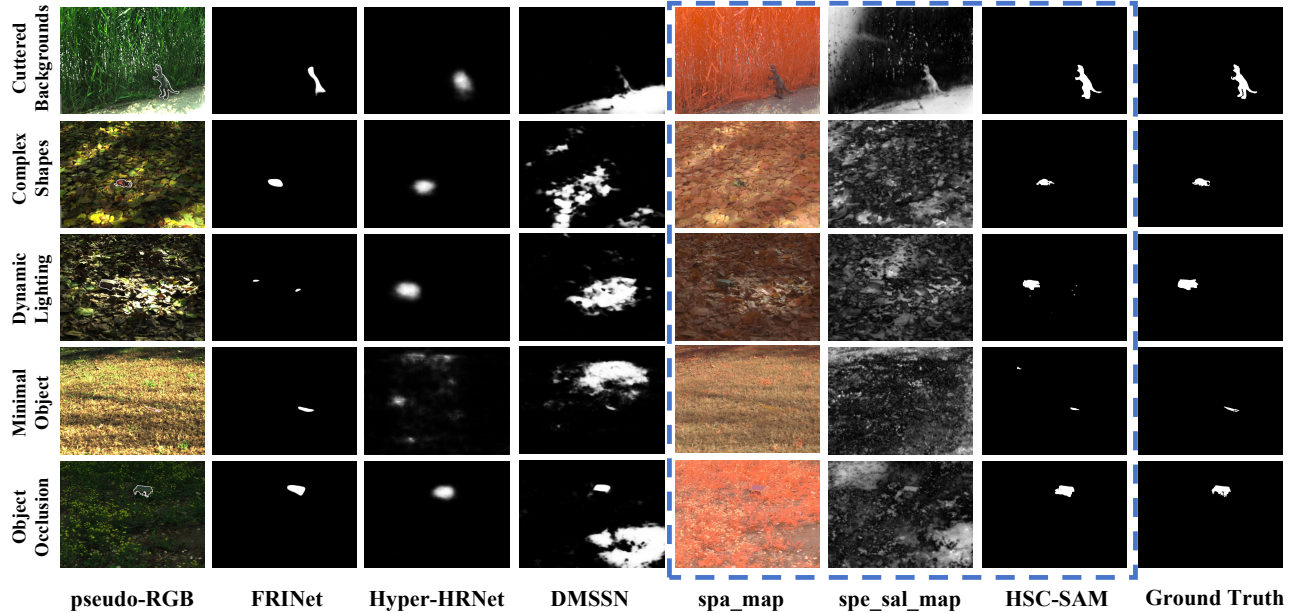


Figure 5: Qualitative results on HyperCOD. HSC-SAM offers clearer contours in challenging camouflage scenarios.

Experiment

Results on HyperCOD

Quantitative Results. We compare HSC-SAM with representative RGB-based camouflaged object detection models on the HyperCOD dataset, as well as several state-of-the-art hyperspectral salient object detection methods for reference. As shown in Tab. 2, RGB-based models built upon SAM’s backbone generally outperform conventional camouflaged object detection models, benefiting from their strong generalization capability. Building on this foundation, our HSC-SAM further bridges the performance gap by effectively leveraging spectral cues, achieving state-of-the-art results across all key metrics. Specifically, HSC-SAM attains the lowest MAE of **0.0017**, outperforming the second-best result by 0.0005, and achieves the highest Adp-F score of **0.681**, representing a 3.1% improvement over the closest competitor, Camoformer. These results demonstrate that incorporating hyperspectral information enables more accurate object localization and finer contour delineation beyond what can be achieved with RGB-based SAM variants alone.

Efficiency Analysis. As shown in Tab. 2, HSC-SAM achieves competitive complexity with **11.7M** parameters and **94.2G** FLOPs, which are substantially lower than those of other SAM-based variants such as SAM2-UNet (216.4M, 128.4G) and high-capacity RGB models like HGINet. These results highlight the effectiveness of HSC-SAM in balancing accuracy and computational efficiency for hyperspectral camouflaged object detection.

Attribute-based Comparison. HSC-SAM demonstrates superior robustness across all challenging scenarios (Tab. 3). The Spectral-Spatial Decomposition Module effectively handles dynamic lighting conditions by revealing spectrally-distinct features, while spectral saliency suppresses background clutter through adaptive token pruning. The framework further refines localization via Spectral-Spatial Complementary Prompting and enhances boundary precision through Fine-Grained Detail Enhancement during training.

Qualitative Results. Fig. 5 illustrates the prediction results of our HSC-SAM compared with other camouflaged object detection methods. In complex and challenging sce-

Light-SAM	SSDM	SSCP	SGTD	FDE	Adp-F \uparrow	E \uparrow
✓	★	✗	✗	✗	0.481	0.782
✓	✓	✗	✗	✗	0.557	0.806
✓	✓	✓	✗	✗	0.596	0.823
✓	✓	✓	✓	✗	0.637	0.831
✓	✓	✓	✓	✓	0.657	0.838
✓	✓	✓	✓	◆	0.681	0.853

Table 4: Ablation study of key components in HSC-SAM. ★: pseudo-RGB input only; ◆: training-only module.

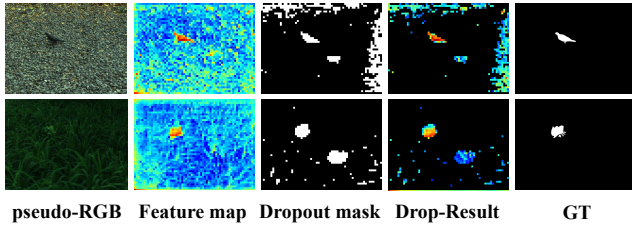


Figure 6: Visualization of the SGTD feature map. SGTD filters low-saliency background features using spectral cues, enabling precise early-stage localization.

narios, HSC-SAM accurately localizes spatially camouflaged yet spectrally salient objects and precisely segments their shapes. Compared to other approaches, HSC-SAM delivers segmentation results with notably sharper and more accurate object boundaries.

Visualization of Background Dropout. To intuitively demonstrate the superiority of HSC-SAM, we visualize its intermediate behavior for qualitative analysis. Fig. 6 illustrates how SGTD leverages the spectral saliency map to suppress background features. By exploiting spectral cues, HSC-SAM effectively highlights target regions and filters out background noise at an early stage, significantly reducing computational overhead.

Ablation Study

Effect of Key Components. We conduct an ablation study to assess the contribution of each component in our HSC-SAM framework, as shown in Tab. 4. The Spectral-Spatial Decomposition Module replaces pseudo-RGB inputs with CIEXYZ-mapped images, achieving a **15.6%** Adp-F improvement by preserving material-specific spectral signatures that are critical for distinguishing true camouflage. Spectral-Spatial Complementary Prompting further increases Adp-F to 0.596 (**7.0%** gain) and E-measure to 0.823 (**2.1%** gain) by establishing joint spectral-spatial reasoning paths. The Spectral-Guided Token Dropout achieves **6.9%** Adp-F improvement and 1.1% E-measure gain through adaptive background suppression. This highlights the module’s effectiveness. Finally, the Fine-Grained Detail Enhancement module enhances boundary precision through training-time multi-scale gradient learning, leaving a **6.9%** Adp-F legacy and producing sharper contours (Fig. 7) even when disabled during inference.

τ	MAE \downarrow	Adp-F \uparrow	E \uparrow	S \uparrow
0.1	0.0018	0.668	0.796	0.783
0.03	0.0020	0.663	0.861	0.788
0.01	0.0017	0.680	0.853	0.802
0.003	0.0020	0.680	0.832	0.789
0.001	0.0019	0.659	0.795	0.782

Table 5: Ablation study on the Spectral-Guided Token Dropout (SGTD) threshold τ .

Setting	Parameters	FPS	Adp-F \uparrow
w/ SGTD	12.09M	19.1	0.639
w/o SGTD	11.74M	20.3 (6.3% \uparrow)	0.681

Table 6: Efficiency analysis of the SGTD module.

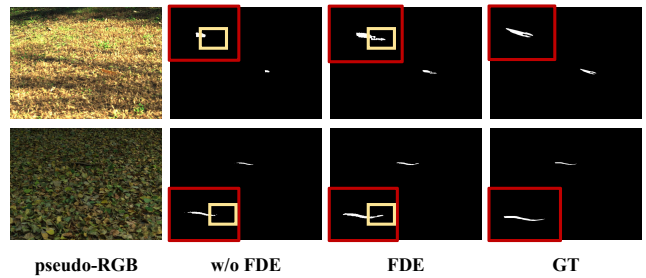


Figure 7: Visual comparison of predictions with and without training-time FDE supervision. FDE-enhanced models recover finer structural details and more accurate boundaries.

SGTD Module Analysis. We present a comprehensive analysis of the SGTD module, covering both its efficiency and the sensitivity of its key hyperparameter. The efficiency analysis (Table 6) demonstrates a favorable trade-off: despite a marginal increase in parameter count and a minor reduction in FPS, the proposed module yields a substantial gain in detection accuracy. Furthermore, we evaluate the robustness of the SGTD threshold τ , which binarizes spectral saliency maps to guide the token dropout. As shown in Tab. 5, $\tau = 0.01$ achieves the optimal balance across all metrics, confirming the module’s stability to minor hyperparameter variations.

Conclusion

In this paper, we introduced HyperCOD, the first large-scale benchmark for hyperspectral camouflaged object detection, featuring 350 high-resolution hyperspectral images across various camouflage scenarios and environmental conditions. We also proposed HSC-SAM, a spectral-aware adaptation of the Segment Anything Model that integrates spectral-spatial decomposition and saliency-guided token filtering to enhance object-relevant features, achieving state-of-the-art performance. We believe that HyperCOD, along with the proposed framework, will drive significant advancements in hyperspectral camouflaged object detection research.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (No. 62571031).

References

- Chen, H.; Wei, P.; Guo, G.; and Gao, S. 2024a. SAM-COD: SAM-guided unified framework for weakly-supervised camouflaged object detection. <https://arxiv.org/abs/2408.10760>.
- Chen, T.; Lu, A.; Zhu, L.; and et al. 2024b. SAM2-Adapter: Evaluating & adapting Segment Anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. <https://arxiv.org/abs/2408.04579>.
- Fan, D.; Ji, G.; Sun, G.; Cheng, M.; Shen, J.; and Shao, L. 2020a. Camouflaged object detection. In *CVPR*, 2777–2787.
- Fan, D.; Ji, G.; Xu, P.; et al. 2023. Advances in deep concealed scene understanding. *Visual Intelligence*, 1(1): 16.
- Fan, D.-P.; Ji, G.-P.; Cheng, M.-M.; and Shao, L. 2022. Concealed Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10): 6024–6042.
- Fan, D.-P.; Zhou, T.; Ji, G.-P.; Zhou, Y.; Chen, G.; Fu, H.; Shen, J.; and Shao, L. 2020b. Inf-Net: Automatic COVID-19 lung infection segmentation from CT images. *IEEE Transactions on Medical Imaging*, 39(8): 2626–2637.
- Foster, D. H.; Nascimento, S. M.; and Amano, K. 2004. Information limits on neural identification of colored surfaces in natural scenes. *Visual Neuroscience*, 21(3): 331–336.
- Galun; Sharon; Basri; and Brandt. 2003. Texture segmentation by multiscale aggregation of filter responses and shape elements. In *ICCV*, 716–723 vol.1.
- Gao, S.; Zhang, P.; Yan, T.; and Lu, H. 2024. Multi-scale and detail-enhanced segment anything model for salient object detection. <https://arxiv.org/abs/2408.04326>.
- Hu, J.; Lin, J.; Yan, J.; and Gong, S. 2024. Leveraging Hallucinations to Reduce Manual Prompt Dependency in Promptable Segmentation. <https://arxiv.org/abs/2408.15205>.
- Imamoglu, N.; Oishi, Y.; Zhang, X.; Ding, G.; Fang, Y.; Kouyama, T.; and Nakamura, R. 2018. Hyperspectral image dataset for benchmarking on salient object detection. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 1–3. IEEE.
- Jia, Q.; Yao, S.; Liu, Y.; and et al. 2022. Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In *CVPR*, 4703–4712.
- Le, T.-N.; Cao, Y.; Nguyen, T.-C.; Le, M.-Q.; Nguyen, K.-D.; Do, T.-T.; Tran, M.-T.; and Nguyen, T. V. 2022. Camouflaged Instance Segmentation In-The-Wild: Dataset, Method, and Benchmark Suite. *IEEE Transactions on Image Processing*, 31: 287–300.
- Le, T.-N.; Nguyen, T. V.; Nie, Z.; Tran, M.-T.; and Sugimoto, A. 2019. Anabran network for camouflaged object segmentation. *Computer Vision and Image Understanding*, 184: 45–56.
- Liu, L.; Wang, R.; Xie, C.; Yang, P.; Wang, F.; Sudirman, S.; and Liu, W. 2019. PestNet: An end-to-end deep learning approach for large-scale multiclass pest detection and classification. *IEEE Access*, 7: 45301–45312.
- Liu, P.; Xu, T.; Chen, H.; Zhou, S.; Qin, H.; and Li, J. 2023. Spectrum-driven Mixed-frequency Network for Hyperspectral Salient Object Detection. *IEEE Transactions on Multimedia*.
- Liu, P.; Xu, T.; Shi, G.; Xu, J.; Chen, H.; and Li, J. 2025. Spectrum-Oriented Point-Supervised Saliency Detector for Hyperspectral Images. *IEEE Transactions on Instrumentation and Measurement*, 74: 1–15.
- Lyu, Y.; Zhang, J.; Dai, Y.; Li, A.; Liu, B.; Barnes, N.; and Fan, D.-P. 2021. Simultaneously Localize, Segment and Rank the Camouflaged Objects. In *CVPR*.
- Pang, Y.; Zhao, X.; Xiang, T.; Zhang, L.; and Lu, H. 2022. Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In *CVPR*, 2160–2170.
- Qin, H.; Xu, T.; Liu, P.; Xu, J.; and Li, J. 2024. DMSSN: Distilled Mixed Spectral-Spatial Network for Hyperspectral Salient Object Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
- Qiu, Y.; Bai, S.; Xu, T.; Liu, P.; Qin, H.; and Li, J. 2025. HSOD-BIT-V2: A new challenging benchmark for hyperspectral salient object detection. <https://arxiv.org/abs/2503.13906>.
- Rao, C. P.; Reddy, A. G.; and Rao, C. B. R. 2020. Camouflaged object detection for machine vision applications. *International Journal of Speech Technology*, 23: 327–335.
- Skurowski, P.; Abdulameer, H.; Błaszczak, J.; Depta, T.; Kornacki, A.; and Koziel, P. 2018. Animal camouflage analysis: Chameleon database. *Unpublished Manuscript*.
- Sun, Y.; Wang, S.; Chen, C.; and et al. 2022. Boundary-guided camouflaged object detection. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, 1335–1341.
- Sun, Y.; Xu, C.; Yang, J.; Xuan, H.; and Luo, L. 2024. Frequency-Spatial Entanglement Learning for Camouflaged Object Detection. <https://arxiv.org/abs/2409.01686>.
- Tankus, A.; and Yeshurun, Y. 1998. Detection of regions of interest and camouflage breaking by direct convexity estimation. In *Proceedings 1998 IEEE Workshop on Visual Surveillance*, 42–48.
- Wang, K.; Bi, H.; and Zhang, Y. 2022. D2C-Net: a dual-branch, dual-guidance and cross-refine network for camouflaged object detection. *IEEE Transactions on Industrial Electronics*, 69(5): 5364–5374.
- Wang, L.; Yang, J.; Zhang, Y.; Wang, F.; and Zheng, F. 2024. Depth-Aware Concealed Crop Detection in Dense Agricultural Scenes. In *CVPR*, 17201–17211.
- Xiong, X.; Wu, Z.; Tan, S.; Li, W.; Tang, F.; Chen, Y.; Li, S.; Ma, J.; and Li, G. 2024. SAM2-UNet: Segment Anything 2 Makes Strong Encoder for Natural and Medical Image Segmentation. *arXiv preprint arXiv:2408.08870*.
- Yao, S.; Sun, H.; Xiang, T.-Z.; Wang, X.; and Cao, X. 2024. Hierarchical Graph Interaction Transformer with Dynamic

Token Clustering for Camouflaged Object Detection. *arXiv preprint arXiv:2408.15020*.

Zheng, Q.; Zheng, L.; Bai, Y.; Liu, H.; Deng, J.; and Li, Y. 2023. Boundary-Aware Network With Two-Stage Partial Decoders for Salient Object Detection in Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–13.