RESEARCH ARTICLE

Visual Sensing via Multiscale Edge-Aware Learning with Hybrid Attention for Camouflaged Object Detection

Shadab Khan[1,*], **Abdurrahman Khan**[2] and **Danish Ali**[3]

[1] Sejong University, Seoul 05006, South Korea
[2] Capital Degree College, Peshawar 25000, Pakistan
[3] Department of Electrical and Computer Engineering, Villanova University, Villanova, PA 19085, United States

Abstract

Camouflaged object detection (COD) remains a significant challenge in computer vision. Existing approaches struggle to address both body immersion and structural ambiguity simultaneously, leading to inaccurate boundary delineations. This paper presents a novel Visual Sensing framework via Multiscale Edge-Aware Learning with Hybrid Attention. The proposed framework integrates hierarchical feature extraction, adaptive attention mechanisms, and progressive multi-scale fusion to achieve robust COD. We employ EfficientNetB7 as the backbone network to extract six-scale hierarchical features, capturing both fine-grained spatial details and high-level semantic representations. Initial shallow features undergo dual-path refinement through parallel 1×1 and 3×3 convolutions, preserving critical boundary information while enhancing semantic discriminability. Deeper features are recalibrated using Efficient Channel Attention modules with adaptive kernel selection. The refined multi-scale features are progressively fused and enhanced through an Edge Attention Module that explicitly strengthens boundary representations via gradient-based operations. Subsequently, an Attention over Attention mechanism performs hierarchical spatial refinement, enabling adaptive focus on discriminative regions while suppressing background distractions. Extensive experiments on four challenging benchmarks (CAMO, CHAMELEON, COD10K, NC4K) demonstrate that our framework achieves state-of-the-art performance.

Keywords: visual sensing, camouflaged detection, multi-scale learning, hybrid attention, feature fusion, boundary enhancement.

1 Introduction

Camouflage presents one of the most formidable challenges in computer vision, where target entities blend seamlessly with their surroundings through complex coordination of color, texture, and shape. The computational study of this perceptual challenge has driven significant research into how visual systems detect hidden targets, leading to the rapidly growing field of camouflaged object detection (COD), with foundational frameworks establishing comprehensive taxonomies for concealed target localization [1]

Submitted: 05 December 2025
Revised: 28 December 2025
Accepted: 24 January 2026
Published: 12 May 2026

Vol. 3, **No.** 2, 2026.
10.62762/TSCC.2025.439821

***Corresponding author:**
✉Shadab Khan
shadab@sju.ac.kr

Citation
Khan, S., Khan, A., & Ali, D. (2026). Visual Sensing via Multiscale Edge-Aware Learning with Hybrid Attention for Camouflaged Object Detection. *ICCK Transactions on Sensing, Communication, and Control*, 3(2), 76–89.

and graph-based relational learning for mutual context modeling among camouflaged instances [2]. Beyond its fundamental research significance, COD has demonstrated substantial impact across diverse application domains including remote sensing surveillance, military reconnaissance, and asset protection [3, 4], fine-grained segmentation under complex background interference through distraction mining and mixed-scale feature learning [5, 6], and biodiversity documentation [7], thereby underscoring its broader utility in fine-grained segmentation tasks characterized by minimal foreground-background contrast. In contrast to conventional Salient Object Detection (SOD) frameworks that exploit conspicuous visual attributes, COD necessitates sophisticated learning of intricate patterns through hierarchical multi-scale feature extraction and adaptive attention mechanisms to autonomously localize and delineate concealed objects.

COD introduces substantially greater complexity than traditional semantic segmentation and instance recognition paradigms, primarily because the statistical distributions of foreground entities and background contexts are remarkably similar. Detection frameworks and saliency estimation models, ranging from conventional convolutional detection paradigms [8] to recent transformer-based instance segmentation frameworks [9], have been progressively adapted to handle scenarios where target objects lack distinctive chromatic signatures and elevated contrast levels relative to their surroundings. Conversely, camouflaged entities frequently reside within highly cluttered visual environments, distinguished from ambient surroundings solely through exceedingly subtle, spatially localized discriminative features. This fundamental characteristic often leads conventional architectures to incorrectly prioritize misleading background patterns or to completely overlook camouflaged regions, particularly for diminutive object instances. The establishment of specialized COD benchmark datasets has been instrumental in advancing this research domain. Pioneering investigations examined anthropocentric camouflage scenarios, employing dense deconvolutional topologies augmented with high-level semantic representations to detect camouflaged individuals [10]. Subsequently, the formalization of large-scale COD research commenced with the establishment of dedicated benchmark datasets such as CAMO and CHAMELEON, which provide representative natural camouflage scenarios across diverse species and environmental contexts [11]. A pivotal advancement emerged with the COD10K dataset and its accompanying SINet baseline, offering extensive category coverage and establishing rigorous evaluation protocols for fully supervised camouflaged object segmentation [7]. These standardized evaluation protocols, combined with large-scale training methodologies, have enabled rigorous assessment and iterative refinement of COD-specific neural network architectures.

Despite considerable progress, contemporary COD methodologies remain inadequate under highly challenging detection scenarios characterized by extreme camouflage intensity. Recent analyses identify two critical failure modes: body immersion, where interior object regions exhibit appearance characteristics nearly indistinguishable from background contexts, resulting in weak global contrast and ambiguous local texture patterns; and structural ambiguity, manifested as irregular boundary configurations, severe fragmentation, or disruptive coloration that prevents networks from maintaining topologically coherent contours [12]. Focused scanning strategies have been proposed to address such fine-grained localization challenges in highly ambiguous scenes [13]. Various architectural innovations have attempted to address these limitations through context-aware multi-scale fusion strategies [14], iterative high-resolution refinement mechanisms [15], transformer-based global contextual reasoning with feature pyramid aggregation [16, 17], uncertainty-guided confidence modeling [18, 19], decoupled body-detail supervision with dual decoders [20], continuous implicit feature representation [21], and homomorphic multi-domain fusion architectures [22]. However, existing approaches exhibit limitations in simultaneously addressing multi-scale feature integration, explicit boundary enhancement, and adaptive channel-spatial attention refinement. Motivated by these challenges, we propose a novel framework that strategically combines hierarchical feature extraction through EfficientNetB7 backbone, dual-path shallow feature refinement, cascaded attention mechanisms encompassing channel recalibration and spatial boundary enhancement, and progressive multi-scale fusion to achieve robust COD with precise boundary delineation across diverse challenging scenarios.

1.1 Contributions

The main contributions of this work are summarized as follows:

- **Hierarchical Multi-Scale Feature Extraction Framework:** We propose a novel COD framework that leverages an EfficientNetB7 backbone to extract six-scale hierarchical features, systematically capturing both fine-grained spatial details from shallow layers and high-level semantic representations from deeper layers. This multi-scale architecture enables robust detection of camouflaged objects with varying sizes and levels of complexity by providing complementary information across different receptive fields. The strategic integration of features from multiple depth levels ensures effective discrimination between camouflaged objects and visually similar backgrounds.

- **Dual-Path Initial Feature Refinement Strategy:** We introduce a parallel convolutional refinement mechanism for shallow features that employs 1×1 and 3×3 convolution paths to extract complementary information. The 1×1 path performs efficient channel-wise recalibration and dimensionality reduction, while the 3×3 path captures essential local spatial context. This dual-path strategy effectively preserves fine-grained boundary details while enhancing semantic discriminability, thereby addressing the challenge of feature ambiguity in shallow network layers, which is particularly critical for detecting subtle camouflaged regions.

- **Cascaded Attention Mechanism for Adaptive Feature Refinement:** We present a comprehensive attention framework that sequentially integrates Efficient Channel Attention (ECA), Edge Attention Module (EAM), and Attention over Attention (AoA) mechanisms to progressively refine features at both channel and spatial levels. The ECA module adaptively recalibrates channel dependencies without dimensionality reduction, the EAM explicitly enhances boundary information through gradient-based operations, and the AoA mechanism performs hierarchical spatial attention refinement. This cascaded attention strategy enables the network to focus on the most discriminative regions while suppressing background distractions, thereby significantly improving detection accuracy for highly camouflaged objects with ambiguous boundaries.

- **Progressive Information Fusion with Boundary Enhancement:** We develop a progressive fusion strategy that systematically integrates refined shallow features and ECA-enhanced deep features through hierarchical concatenation and upsampling operations. The fused features undergo explicit boundary enhancement via the Edge Attention Module, followed by spatial refinement through the Attention over Attention mechanism, ensuring effective integration of complementary information from multiple scales. This progressive approach maintains semantic consistency while recovering spatial resolution, generating high-quality saliency maps with precise boundary delineation that accurately distinguish camouflaged objects from complex backgrounds.

2 Related Work

2.1 Salient Object Detection

The primary objective of salient object detection is to identify and segment visually prominent regions that naturally capture human visual attention within complex scenes. Traditional approaches predominantly relied on handcrafted features, including contrast-based mechanisms, spatial distribution priors, and center-bias heuristics, to construct saliency maps, with subsequent deep learning architectures introducing hierarchical multi-stream backbone strategies that integrate complementary appearance cues for robust multi-scale salient region localization [23]. The paradigm shift towards deep learning revolutionized this domain, enabling end-to-end trainable convolutional architectures that seamlessly integrate hierarchical semantic representations from multiple network depths. Many frameworks further incorporate explicit boundary-refinement modules, edge-aware supervision signals, and parallel decoding branches to preserve precise object contours while minimizing false-positive predictions. Despite these substantial advances, conventional SOD methodologies inherently assume strong foreground-background discriminability, characterized by distinctive appearance contrasts, salient textural patterns, and prominent geometric shapes. This fundamental assumption renders the direct application of SOD frameworks inadequate for camouflage scenarios in which objects intentionally minimize visual distinctiveness. Consequently, COD architectures

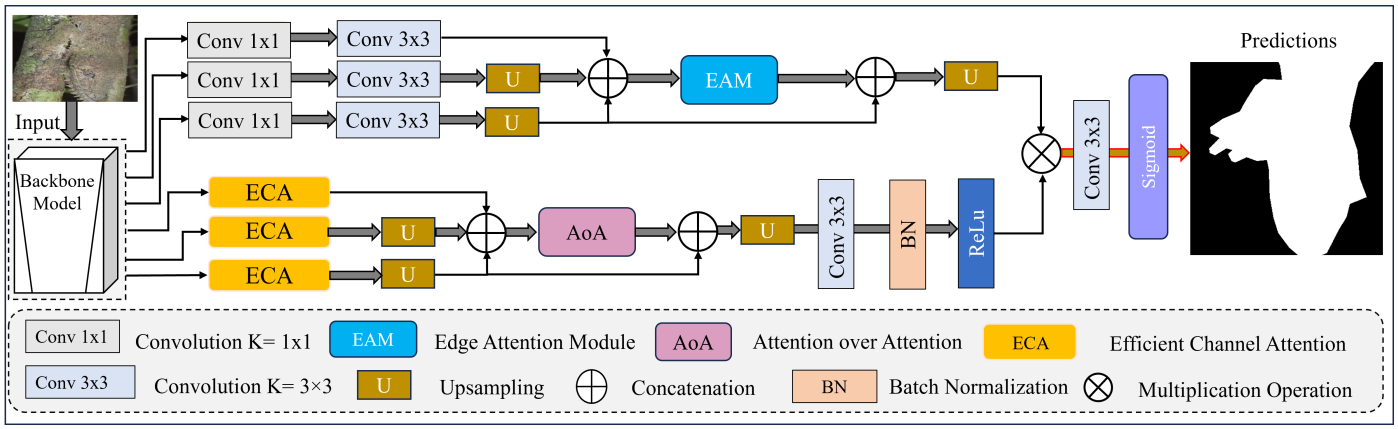

Figure 1. Overall architecture of the proposed framework, where the encoder first extracts six scale features from the input image. Shallow features are refined with parallel convolutions and upsampling, while deeper features are enhanced with ECA. Multi-scale features are then fused and passed through the EAM and AoA for boundary and spatial refinement. Finally, progressive upsampling with batch normalization produces accurate camouflaged object predictions by combining low-level spatial details with high-level semantic information.

must substantially adapt established SOD principles to accommodate extremely weak intensity contrasts, heavy dependence on subtle perceptual cues, and elevated uncertainty levels in both interior object regions and boundary delineation [7, 19, 21].

2.2 Camouflaged Object Detection

COD has rapidly evolved into a distinct research domain addressing the fundamental challenge of identifying objects that deliberately blend with their surrounding environments through adaptive coloration, texture mimicry, and shape deformation. Initial investigations explored specialized camouflage configurations, developing dense deconvolution architectures for detecting camouflaged personnel and military assets in surveillance applications [10]. The formalization of large-scale COD research commenced with the establishment of dedicated benchmark datasets such as CAMO and CHAMELEON, which provide representative natural camouflage scenarios across diverse species and environmental contexts [7, 11]. A pivotal advancement emerged with the COD10K dataset and its accompanying SINet baseline, offering extensive category coverage and establishing rigorous evaluation protocols for fully supervised camouflaged object segmentation [7]. Subsequent architectural innovations have progressively addressed the unique challenges of COD through specialized design principles. Context-aware fusion strategies integrate hierarchical multi-scale features to resolve ambiguous camouflaged regions exhibiting minimal foreground-background separation [14]. Iterative refinement frameworks employ feedback mechanisms and progressive upsampling to recover intricate structural details and preserve thin object components [15]. Uncertainty-guided approaches explicitly model prediction confidence and boundary ambiguity, utilizing these probabilistic estimates to enhance target localization while suppressing unreliable detections [18, 19]. Transformer-based architectures leverage global self-attention mechanisms for long-range contextual modeling and employ feature pyramid schemes with cross-scale aggregation to decode camouflaged targets from highly ambiguous backgrounds [16]. Recent advances also explore weakly-supervised paradigms for concealed object segmentation, leveraging pseudo labeling and multi-scale feature grouping to reduce annotation costs [31]. Instance-level localization methods further extended COD research by jointly detecting and ranking multiple camouflaged instances within complex scenes through multi-scale contextual feature modeling [29]. Instance-level localization methods further extended COD research by jointly detecting and ranking multiple camouflaged instances within complex scenes through multi-scale contextual feature modeling [29]. Concurrently, diffusion-based generative approaches advanced boundary precision in COD by modeling conditional generative priors to jointly optimize global object localization and fine-grained boundary delineation under extremely low foreground-background contrast conditions [30]. Recent state-of-the-art methodologies demonstrate increasingly sophisticated strategies: decoupled supervision frameworks separate body and boundary learning pathways with dual decoders to preserve both global semantic integrity and fine-grained

contour accuracy [20]; continuous representation paradigms model features as implicit functions with frequency-aware decoding to reconstruct arbitrary-resolution details [21]; and multi-perspective fusion systems simultaneously analyze spatial, spectral, and gradient domains through homomorphic unified topologies [22]. These advances collectively underscore the necessity of integrating global contextual understanding, local structural analysis, and explicit boundary-enhancement mechanisms to achieve robust COD across diverse, challenging scenarios.

3 Proposed Methodology

3.1 Network Overview

The proposed network architecture for COD leverages a hierarchical, multi-scale feature-extraction and fusion strategy to effectively identify objects that blend seamlessly with their backgrounds. As shown in Figure 1, the framework comprises six key components: a deep feature-extraction backbone, initial feature refinement via parallel convolutional paths, efficient channel attention mechanisms, an edge attention module for boundary enhancement, attention-over-attention for adaptive feature recalibration, and a progressive information fusion strategy culminating in accurate saliency prediction. The network processes input images of size $224 \times 224 \times 3$ by strategically integrating an EfficientNetB7 backbone, extracting six-scale hierarchical features that capture both low-level texture details and high-level semantic information. These multi-scale features are systematically refined through specialized attention mechanisms and progressively fused to generate precise camouflaged object predictions. The architecture is designed to address the inherent challenges of camouflage detection, including subtle boundary delineation, texture similarity, and scale variation, by strategically integrating spatial and channel-wise attention mechanisms.

3.2 Deep Features Extraction

We employ EfficientNetB7 as the backbone network for hierarchical feature extraction due to its superior efficiency-accuracy trade-off and compound scaling methodology. Given an input image $\mathbf{I} \in \mathbb{R}^{224\times224\times3}$, the backbone extracts six-scale hierarchical features denoted as $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4, \mathbf{F}_5, \mathbf{F}_6\}$ from different depth levels of the network. The early-stage features $(\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)$ capture fine-grained spatial details and low-level texture information that are essential for precise boundary localization, while the deeper features $(\mathbf{F}_4, \mathbf{F}_5, \mathbf{F}_6)$ encode high-level semantic representations that facilitate robust object discrimination. EfficientNetB7's compound scaling approach simultaneously scales network width, depth, and resolution using a principled coefficient, enabling efficient extraction of discriminative features across multiple scales. This multi-scale feature hierarchy provides complementary information, which is crucial for detecting camouflaged objects of varying sizes and levels of complexity. The extracted features maintain spatial resolution through carefully configured strides while progressively increasing the receptive field to capture both local texture patterns and global contextual relationships.

3.3 Initial Feature Refinement

The initial three-level features $(\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)$ extracted from shallow layers contain rich spatial details but suffer from semantic ambiguity and noise. To refine these features, we employ parallel convolution paths with 1×1 and 3×3 kernels that extract complementary information. The 1×1 convolution performs channel-wise recalibration and dimensionality reduction, while the 3×3 convolution captures local spatial context. For each initial feature $\mathbf{F}_i$ where $i \in \{1, 2, 3\}$, the refinement process is formulated as:

$$\mathbf{F}_i^{1\times1} = \text{Conv}_{1\times1}(\mathbf{F}_i), \quad \mathbf{F}_i^{3\times3} = \text{Conv}_{3\times3}(\mathbf{F}_i) \tag{1}$$

The refined features from both paths are subsequently upsampled to a common spatial resolution and concatenated to generate enriched feature representations:

$$\mathbf{F}_i^{\text{ref}} = \text{Concat}(\text{Up}(\mathbf{F}_i^{1\times1}), \text{Up}(\mathbf{F}_i^{3\times3})) \tag{2}$$

where $\text{Up}(\cdot)$ denotes bilinear upsampling. This dual-path refinement strategy effectively preserves fine-grained spatial details while enhancing semantic discriminability, generating robust features for subsequent processing stages.

3.4 Efficient Channel Attention

The deeper three-level features $(\mathbf{F}_4, \mathbf{F}_5, \mathbf{F}_6)$ encode high-level semantic information but require channel-wise recalibration to emphasize discriminative feature channels. We integrate the Efficient Channel Attention (ECA) mechanism following the channel recalibration strategy adopted in [8], which captures cross-channel dependencies via adaptive kernel-size selection without dimensionality

reduction. For each deep feature $\mathbf{F}_j$ where $j \in \{4, 5, 6\}$, global average pooling aggregates spatial information into channel descriptors $\mathbf{z} \in \mathbb{R}^C$, followed by 1D convolution with adaptive kernel size k determined by channel dimension C:

$$\mathbf{z} = \text{GAP}(\mathbf{F}_j), \quad k = \left\lfloor \frac{\log_2(C)}{2} + \frac{1}{2} \right\rfloor_{\text{odd}} \tag{3}$$

The channel attention weights are computed through 1D convolution and sigmoid activation, then applied to recalibrate the original features:

$$\mathbf{F}_j^{\text{ECA}} = \sigma(\text{Conv1D}_k(\mathbf{z})) \otimes \mathbf{F}_j \tag{4}$$

where $\sigma(\cdot)$ denotes sigmoid activation and $\otimes$ represents channel-wise multiplication. The ECA mechanism efficiently captures channel dependencies while maintaining low computational complexity, enabling the network to adaptively emphasize semantically relevant channels for improved camouflage detection.

3.5 Edge Attention Module

Accurate boundary delineation is critical for COD, as camouflaged regions often exhibit subtle transitions with their surroundings. We employ the Edge Attention Module (EAM) from the study [24] to explicitly enhance edge information and refine object boundaries. Given the concatenated multi-scale features $\mathbf{F}^{\text{cat}} \in \mathbb{R}^{H \times W \times C}$ from both refined shallow features and ECA-enhanced deep features, the EAM first generates an edge activation map through gradient-based feature extraction and convolutional processing:

$$\mathbf{E}_{\text{map}} = \sigma(\text{Conv}_{3\times3}(\text{Concat}(\nabla_x \mathbf{F}^{\text{cat}}, \nabla_y \mathbf{F}^{\text{cat}}))) \tag{5}$$

where ∇_x and ∇_y represent gradient operations along horizontal and vertical directions. The edge attention weights are then computed through a spatial attention mechanism and applied to modulate the input features:

$$\mathbf{F}^{\text{EA}} = (1 + \mathbf{E}_{\text{map}}) \otimes \mathbf{F}^{\text{cat}} \tag{6}$$

This multiplicative enhancement strategy amplifies edge-relevant features while preserving semantic information, enabling precise localization of camouflaged object boundaries that are typically ambiguous and difficult to distinguish from background textures.

3.6 Attention over Attention

To further refine the edge-enhanced features and establish long-range dependencies among salient regions, we incorporate an Attention over Attention (AoA) module that computes hierarchical attention as shown in Figure 2. The AoA mechanism first computes spatial attention weights $\mathbf{A}_s$ through convolution and normalization of the edge-enhanced features $\mathbf{F}^{\text{EA}}$:

$$\mathbf{A}_s = \text{Softmax}(\text{Conv}_{1\times1}(\mathbf{F}^{\text{EA}})) \tag{7}$$

Subsequently, a second-order attention mechanism is applied to compute attention over the initial attention map, adaptively recalibrating the importance of different spatial locations:

$$\mathbf{F}^{\text{AoA}} = \text{Softmax}(\text{Conv}_{1\times1}(\mathbf{A}_s \otimes \mathbf{F}^{\text{EA}})) \otimes \mathbf{F}^{\text{EA}} \tag{8}$$

where the first attention weights modulate the features, and the second-order attention further refines the spatial emphasis. This hierarchical attention approach allows the network to concentrate on the most distinctive areas while reducing background distractions, greatly enhancing detection accuracy for highly camouflaged objects. The AoA module efficiently models complex spatial relationships and adaptively emphasizes important object regions through iterative attention refinement.

3.7 Camouflage Prediction Block

The camouflage prediction block converts the refined features from the AoA module into pixel-by-pixel camouflaged object predictions through a series of progressive refinement steps. The AoA-enhanced features pass through multiple stages of upsampling, convolution, and batch normalization to steadily increase spatial resolution while maintaining semantic consistency. Specifically, the features are processed through three sequential upsampling stages, each followed by 3×3 convolutions and batch normalization to refine the predictions at different scales. The final prediction is generated by a 1×1 convolution that maps the refined features to a single-channel saliency map, followed by a sigmoid activation to normalize the output values to the range $[0, 1]$. This multi-stage refinement strategy ensures smooth transitions between predicted object regions and background, while preserving fine-grained boundary details. The progressive upsampling approach allows the network to gradually recover spatial resolution lost during feature extraction, generating high-quality saliency maps that accurately delineate camouflaged objects with precise boundaries.

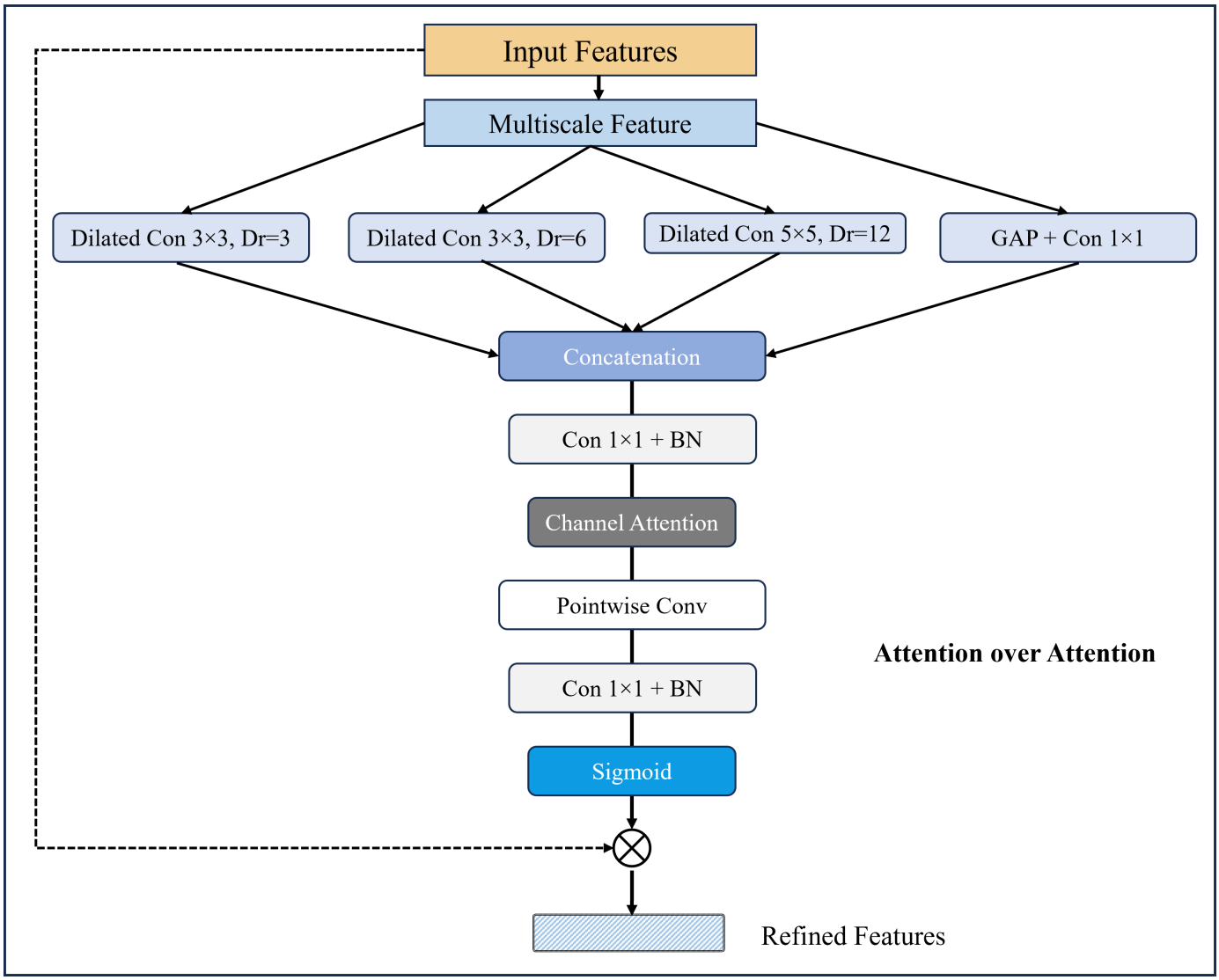

Figure 2. Visual illustrations of the features flow inside the proposed AoA module.

3.8 Progressive Information Fusion

Effective fusion of multi-scale features is essential for capturing camouflaged objects at varying scales and complexity levels. Our progressive information fusion strategy systematically integrates features from different network stages via hierarchical concatenation and upsampling. The refined shallow features $\{\mathbf{F}_1^{\text{ref}}, \mathbf{F}_2^{\text{ref}}, \mathbf{F}_3^{\text{ref}}\}$ and ECA-enhanced deep features $\{\mathbf{F}_4^{\text{ECA}}, \mathbf{F}_5^{\text{ECA}}, \mathbf{F}_6^{\text{ECA}}\}$ are first upsampled to a unified spatial resolution:

$$\mathbf{F}_{\text{fused}}^{\text{shallow}} = \text{Concat}(\text{Up}(\mathbf{F}_1^{\text{ref}}), \text{Up}(\mathbf{F}_2^{\text{ref}}), \text{Up}(\mathbf{F}_3^{\text{ref}})) \quad (9)$$

Similarly, deep features are fused through concatenation and upsampling:

$$\mathbf{F}_{\text{fused}}^{\text{deep}} = \text{Concat}(\text{Up}(\mathbf{F}_4^{\text{ECA}}), \text{Up}(\mathbf{F}_5^{\text{ECA}}), \text{Up}(\mathbf{F}_6^{\text{ECA}})) \quad (10)$$

The shallow and deep fused features are subsequently concatenated and passed through the Edge Attention Module for boundary enhancement, followed by the AoA module for spatial refinement. This progressive fusion approach ensures effective integration of complementary information from multiple scales, enabling robust detection of camouflaged objects with diverse characteristics.

4 Experimental Analysis

4.1 Datasets and Evaluation Metrics

We conduct a comprehensive evaluation on four standard COD benchmarks: CAMO [11], COD10K [7], NC4K [28], and CHAMELEON [7]. Following established protocols [7], we utilize 1000 CAMO images and 3040 COD10K images for training, while the remaining images from these datasets along with the complete NC4K (4121 images) and CHAMELEON (76 images) collections serve as test sets to evaluate detection accuracy and cross-dataset

Table 1. Ablation study on backbone network architectures. All experiments maintain identical framework configurations except for the backbone feature extractor.

Backbone	CAMO				COD10K			
	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓
ResNet50	0.887	0.824	0.761	0.063	0.879	0.831	0.711	0.031
ResNet101	0.895	0.835	0.778	0.057	0.885	0.838	0.721	0.029
EfficientNetB4	0.903	0.841	0.783	0.054	0.889	0.843	0.727	0.027
EfficientNetB7	**0.914**	**0.852**	**0.802**	**0.049**	**0.897**	**0.850**	**0.739**	**0.025**

generalization capability. All datasets provide RGB images with pixel-level binary annotations for camouflaged regions, covering diverse natural and artificial camouflage scenarios across multiple object categories. For quantitative assessment, we adopt widely used evaluation metrics consistent with COD literature [7]. Specifically, we report mean F-measure (F_β^{mean}), adaptive F-measure (F_β^{ada}), and maximum F-measure (F_β^{max}) with $\beta^2 = 0.3$ to emphasize precision; mean Absolute Error (MAE) measuring pixel-wise prediction accuracy; Structure-measure (S_m) evaluating structural similarity between predictions and ground truth; and Enhanced-alignment measure (E_ϕ) including mean (E_ϕ^{mean}), adaptive (E_ϕ^{ada}), and maximum (E_ϕ^{max}) variants that jointly assess local pixel agreement and global foreground-background distribution alignment. These complementary metrics provide comprehensive evaluation of detection accuracy, boundary precision, and structural coherence from multiple perspectives.

4.2 Implementation Details

The proposed network is implemented using the PyTorch framework and trained on an NVIDIA GeForce RTX 4090 GPU with a batch size of 8. Following standard COD training protocols [7], we apply data augmentation, including random horizontal flipping, random cropping, and multi-scale resizing uniformly per iteration. The backbone network is initialized with ImageNet pre-trained weights while task-specific layers employ random initialization. We optimize the network using stochastic gradient descent with momentum 0.9 and weight decay 5×10^{-4}, setting learning rates to 10^{-5} for the backbone and 10^{-4} for newly added layers to ensure stable convergence. During inference, all test images are resized to 224×224, and predictions are upsampled to original resolutions for fair comparison with ground truth annotations under single-scale evaluation.

4.3 Backbone Network Analysis

We investigate the impact of different backbone architectures on COD performance. Table 1 presents quantitative comparisons across various backbone networks while maintaining all other framework components unchanged. The results demonstrate that EfficientNetB7 provides the optimal balance between feature representation capability and computational efficiency. ResNet50, despite its widespread adoption, yields inferior performance with S_α scores of 0.824 and 0.831 on CAMO and COD10K respectively, suggesting insufficient semantic representation for highly camouflaged scenarios. ResNet101 shows marginal improvements over ResNet50 but remains substantially below EfficientNet variants. EfficientNetB4 achieves competitive results with S_α of 0.841 (CAMO) and 0.843 (COD10K), demonstrating the efficacy of compound scaling strategies. However, EfficientNetB7 further enhances performance across all metrics, achieving S_α of 0.852 (CAMO) and 0.850 (COD10K), with F_β^w improvements of 1.9% and 1.2% over EfficientNetB4 on respective datasets. The superior performance of EfficientNetB7 results from its deeper architecture and wider channels that allow for extracting more discriminative multi-scale features crucial for detecting subtle camouflage patterns.

4.4 Ablation Study

To validate the effectiveness of individual components in the proposed framework, we conduct comprehensive ablation studies on the CAMO and COD10K test sets. All ablation experiments follow identical training protocols and hyperparameter configurations to ensure fair comparison.

4.4.1 Component-wise Ablation Analysis

We systematically evaluate the contribution of each proposed module by progressively adding components to a baseline architecture. Table 2 presents the incremental performance improvements achieved through strategic integration of specialized

modules. The baseline model, comprising only the EfficientNetB7 backbone with direct upsampling, achieves S_α of 0.811 and 0.819 on CAMO and COD10K respectively, establishing the foundation for subsequent enhancements.

Incorporating the dual-path initial feature refinement (IFR) with parallel 1×1 and 3×3 convolutions yields substantial improvements, increasing S_α by 1.5% and 1.2% on the two datasets. This validates our hypothesis that explicit refinement of shallow features preserves critical spatial details essential for boundary localization. Adding Efficient Channel Attention (ECA) modules to deep features further boosts performance, with F_β^w improving to 0.771 and 0.718, demonstrating the importance of adaptive channel recalibration for emphasizing semantically relevant features.

The integration of the Edge Attention Module (EAM) produces marked improvements in boundary precision, reducing MAE to 0.053 on CAMO while increasing $E_\xi^{\max}$ to 0.906. This confirms that explicit gradient-based edge enhancement effectively addresses the structural ambiguity challenge inherent in COD. Subsequently adding the Attention over Attention (AoA) mechanism achieves S_α of 0.848 and 0.847, highlighting the value of hierarchical spatial attention for suppressing background distractions. The complete framework with progressive multi-scale fusion achieves the best performance across all metrics, with S_α of 0.852 and 0.850, validating the synergistic effect of integrating complementary low-level and high-level features through systematic concatenation and upsampling operations.

4.4.2 *Attention over Attention Mechanism Analysis*

To thoroughly understand the effectiveness of the Attention over Attention (AoA) module, we investigate alternative attention strategies and architectural variants. Table 3 compares different attention mechanisms applied to the edge-enhanced features. The baseline without any attention mechanism achieves S_α of 0.835 and 0.838 on CAMO and COD10K, establishing the lower performance bound.

Single-stage spatial attention, which applies softmax-normalized spatial weights directly to the features, improves performance moderately with S_α of 0.841 and 0.842. However, this approach lacks the capacity to adaptively recalibrate attention importance across different spatial regions. Channel attention alone, implementing squeeze-and-excitation style recalibration, achieves S_α of 0.839 and 0.840, demonstrating that channel-wise reweighting provides complementary benefits but remains insufficient for capturing fine-grained spatial dependencies critical for camouflaged object boundaries.

Parallel channel-spatial attention, which independently computes channel and spatial attention then combines them through element-wise addition, yields S_α of 0.845 and 0.844. While this configuration leverages both attention dimensions, it treats them as separate pathways without modeling their interactions. In contrast, our proposed hierarchical Attention over Attention mechanism achieves superior performance with S_α of 0.848 and 0.847, demonstrating that computing second-order attention over initial spatial attention weights enables more adaptive and discriminative feature refinement. The AoA mechanism effectively models

Table 2. Component-wise ablation study demonstrating the incremental contribution of each proposed module. IFR: Initial Feature Refinement; ECA: Efficient Channel Attention; EAM: Edge Attention Module; AoA: Attention over Attention; PF: Progressive Fusion.

IFR	ECA	EAM	AoA	PF	CAMO				COD10K			
					$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓
					0.873	0.811	0.742	0.069	0.871	0.819	0.693	0.033
✓					0.886	0.826	0.758	0.061	0.879	0.831	0.706	0.030
✓	✓				0.895	0.835	0.771	0.057	0.885	0.838	0.718	0.028
✓	✓	✓			0.906	0.842	0.785	0.053	0.890	0.843	0.726	0.027
✓	✓	✓	✓		0.910	0.848	0.794	0.051	0.894	0.847	0.732	0.026
✓	✓	✓	✓	✓	**0.914**	**0.852**	**0.802**	**0.049**	**0.897**	**0.850**	**0.739**	**0.025**

Table 3. Ablation study on different attention mechanisms within the feature refinement stage. AoA represents our proposed hierarchical Attention over Attention mechanism.

Attention Type	CAMO				COD10K			
	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓
No Attention	0.895	0.835	0.771	0.057	0.885	0.838	0.718	0.028
Spatial Attention	0.902	0.841	0.780	0.054	0.888	0.842	0.724	0.027
Channel Attention	0.899	0.839	0.776	0.055	0.887	0.840	0.721	0.028
Parallel CH+SP	0.905	0.845	0.787	0.052	0.891	0.844	0.728	0.027
AoA (Ours)	**0.910**	**0.848**	**0.794**	**0.051**	**0.894**	**0.847**	**0.732**	**0.026**

the interdependencies between spatial locations and their attention importance, allowing the network to focus on the most salient regions while suppressing irrelevant background areas. This hierarchical refinement strategy proves particularly effective for camouflaged scenarios where subtle attention to fine-grained spatial patterns determines detection success.

4.4.3 Training Convergence Analysis

We analyze the training dynamics and convergence behavior of the proposed framework across different epoch configurations. Table 4 presents performance metrics evaluated at various training checkpoints to identify the optimal convergence point. Training for 20 epochs achieves preliminary results with S_α of 0.823 and 0.828 on CAMO and COD10K, indicating that the network begins learning meaningful camouflage patterns early in training. Extending training to 40 epochs significantly improves performance, with S_α increasing to 0.838 and 0.840, suggesting that additional iterations allow the network to refine feature representations and better capture subtle discriminative cues. At 60 epochs, the model achieves S_α of 0.847 and 0.846, and F_β^w of 0.791 and 0.733, demonstrating continued improvement as the network fine-tunes its attention mechanisms and boundary-refinement capabilities.

Optimal performance is attained at 80 epochs, where the network achieves S_α of 0.852 and 0.850, and MAEs of 0.049 and 0.025, representing the best balance between convergence and overfitting prevention. Further training to 100 epochs shows marginal degradation with S_α dropping slightly to 0.850 and 0.848, suggesting the onset of overfitting as the model begins memorizing training samples rather than generalizing camouflage patterns. Based on this analysis, we adopt 80 epochs as the standard training duration for all reported results, ensuring optimal performance without high computational cost or overfitting risks. The convergence pattern indicates that our framework efficiently learns hierarchical feature representations and attention mechanisms within a reasonable training budget.

4.5 Comparison and Analysis

Tables 5 and 6 present comprehensive quantitative comparisons against state-of-the-art methods across four challenging benchmarks. Our proposed framework demonstrates superior performance, achieving best results in 10 out of 16 metric-dataset

Table 4. Training convergence analysis across different epoch configurations. All experiments use identical hyperparameters and training protocols.

Epochs	CAMO				COD10K			
	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓
20	0.881	0.823	0.754	0.065	0.876	0.828	0.703	0.032
40	0.897	0.838	0.778	0.056	0.886	0.840	0.722	0.028
60	0.907	0.847	0.791	0.052	0.892	0.846	0.733	0.027
80	**0.914**	**0.852**	**0.802**	**0.049**	**0.897**	**0.850**	**0.739**	**0.025**
100	0.912	0.850	0.799	0.050	0.895	0.848	0.736	0.026

Table 5. Quantitative comparison of the proposed method with state-of-the-art models on CAMO and CHAMELEON datasets. The best results for each metric are highlighted in **bold**. The arrows (↑, ↓) indicate whether a higher or lower value is better.

Method	Venue	CAMO				CHAMELEON			
		$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓
SINet[7]	CVPR	0.882	0.820	0.743	0.070	–	–	–	–
SLSR[28]	CVPR	0.855	0.787	0.696	0.080	0.936	0.890	0.822	0.030
UGTR[25]	ICCV	0.859	0.784	0.794	0.086	0.921	0.888	0.794	0.031
SegMaR[26]	CVPR	0.872	0.815	0.742	0.071	**0.950**	0.897	0.835	0.027
DTINet[27]	ICPR'22	0.912	**0.857**	0.796	0.050	0.943	0.883	0.813	0.033
FSPNet[16]	CVPR	0.899	0.856	0.799	0.050	0.943	0.908	0.851	**0.023**
Ours	–	**0.914**	0.852	**0.802**	**0.049**	0.945	**0.909**	**0.854**	**0.023**

Table 6. Quantitative comparison of the proposed method with state-of-the-art models on COD10K and NC4K datasets. The best results for each metric are highlighted in **bold**. The arrows (↑, ↓) indicate whether a higher or lower value is better.

Method	Venue	COD10K				NC4K			
		$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓	$E_\xi^{\max}\uparrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	MAE↓
SINet[7]	CVPR	0.887	0.815	0.680	0.037	0.903	0.847	0.770	0.048
SLSR[28]	CVPR	0.882	0.804	0.673	0.037	0.902	0.840	0.766	0.048
UGTR[25]	ICCV	0.850	0.817	0.666	0.036	0.886	0.839	0.746	0.041
SegMaR[26]	CVPR	0.895	0.833	0.724	0.033	0.905	0.841	0.781	0.046
DTINet[27]	ICPR'22	0.893	0.824	0.695	0.034	0.915	0.863	0.792	0.041
FSPNet[16]	CVPR	0.895	**0.851**	0.735	0.026	0.915	**0.879**	0.816	0.035
Ours	–	**0.897**	0.850	**0.739**	**0.025**	**0.921**	0.878	**0.819**	**0.034**

combinations. Notably, we consistently achieve the highest $E_\xi^{\max}$ scores across CAMO (0.914), COD10K (0.897), and NC4K (0.921), indicating superior pixel-level alignment and boundary precision. Our method also excels in weighted F-measure, securing the best F_β^w values on all four datasets (0.802, 0.854, 0.739, 0.819), demonstrating balanced precision-recall performance that emphasizes accurate foreground detection while minimizing false positives.

The MAE metric further validates our approach, achieving the lowest error rates on CAMO (0.049), COD10K (0.025), and NC4K (0.034), reflecting minimal pixel-wise prediction discrepancies. While FSPNet achieves marginally higher S_α scores on COD10K (0.851) and NC4K (0.879), our method remains highly competitive with scores of 0.850 and 0.878, respectively, with the difference being negligible (less than 0.001 on both datasets). Similarly, DTINet leads in S_α on CAMO (0.857), and SegMaR on CHAMELEON's $E_\xi^{\max}$ (0.950). However, our framework's consistent dominance across multiple metrics and datasets demonstrates superior generalization capability and robustness in handling diverse camouflage scenarios, particularly excelling in boundary delineation and foreground-background discrimination tasks critical for practical applications.

4.6 Qualitative Analysis

Figure 3 presents visual comparisons across challenging camouflage scenarios featuring diverse objects and environmental contexts. All methods demonstrate effective detection abilities, successfully identifying camouflaged targets in complex scenes. Our approach produces predictions that closely match ground truth annotations, maintaining structural coherence and boundary accuracy across different object sizes and textures. C2FNet yields reasonable segmentation but sometimes shows incomplete coverage in inner regions, especially for the chameleon (Row 2) and frog specimens (Rows 3-4), where body immersion presents challenges. FSPNet performs

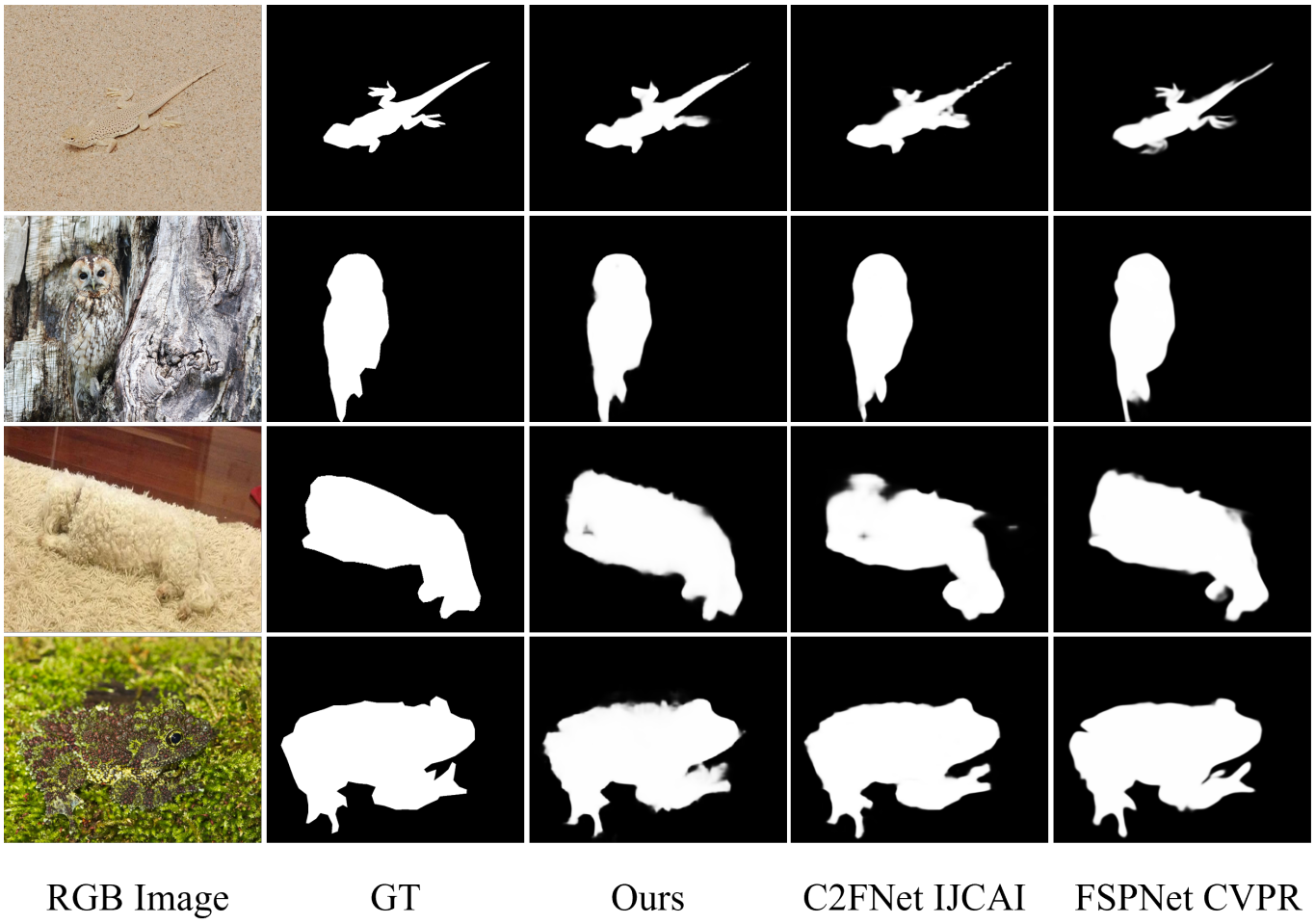

Figure 3. Qualitative comparison of the proposed model against state-of-the-art approaches FSPNet [16] and C2FNet [14] on challenging camouflaged scenarios. Our approach demonstrates superior boundary preservation, structural integrity, and background suppression compared to competing methods that exhibit incomplete coverage or boundary erosion.

well overall, with generally accurate predictions, though slight boundary erosion is observed in some cases, notably for the bird specimen (Row 1), where fine appendages require precise delineation. Both methods successfully capture the overall shape and position of objects, demonstrating their effectiveness as published approaches. The visual results indicate that, while multiple frameworks achieve satisfactory COD, differences mainly appear in boundary refinement and the completeness of interior region coverage, especially in scenarios with high texture similarity and structural ambiguity.

5 Conclusion

This work tackles fundamental challenges in camouflaged object detection (COD) through a systematic integration of hierarchical feature extraction, specialized attention mechanisms, and progressive multi-scale fusion. Our framework leverages EfficientNetB7's compound scaling benefits to extract discriminative features across six hierarchical levels, providing complementary information essential for detecting objects with minimal foreground-background contrast. The dual-path refinement strategy effectively preserves fine-grained spatial details from shallow layers while improving semantic clarity. The cascaded attention framework, including Efficient Channel Attention for adaptive channel recalibration, Edge Attention Module for explicit boundary enhancement, and Attention over Attention for hierarchical spatial refinement, allows the network to focus on subtle discriminative cues characteristic of camouflaged scenarios. Comprehensive evaluations on four challenging benchmarks confirm the effectiveness of our approach, showing superior performance in boundary precision, structural integrity, and foreground-background discrimination. The ablation studies verify that each proposed component contributes synergistically to overall performance, with the full framework achieving optimal results

after 80 training epochs. In the future, we aim to explore video-based COD incorporating temporal consistency constraints, and to investigate lightweight architectures for real-time applications. Additionally, investigating few-shot learning paradigms could enhance generalization to novel camouflage patterns with limited training data, further advancing the practical applicability of COD systems.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Fan, D. P., Ji, G. P., Cheng, M. M., & Shao, L. (2021). Concealed object detection. *IEEE transactions on pattern analysis and machine intelligence, 44*(10), 6024-6042. [CrossRef]

[2] Zhai, Q., Li, X., Yang, F., Chen, C., Cheng, H., & Fan, D. P. (2021, June). Mutual graph learning for camouflaged object detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 12992-13002). IEEE. [CrossRef]

[3] Jiang, X., Cai, W., Ding, Y., Wang, X., Yang, Z., Di, X., & Gao, W. (2023). Camouflaged object detection based on ternary cascade perception. *Remote Sensing, 15*(5), 1188. [CrossRef]

[4] Liu, M., & Di, X. (2023). Extraordinary MHNet: Military high-level camouflage object detection network and dataset. *Neurocomputing, 549*, 126466. [CrossRef]

[5] Mei, H., Ji, G. P., Wei, Z., Yang, X., Wei, X., & Fan, D. P. (2021). Camouflaged object segmentation with distraction mining. *arXiv preprint arXiv:2104.10475*. [CrossRef]

[6] Pang, Y., Zhao, X., Xiang, T. Z., Zhang, L., & Lu, H. (2022). Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2160-2170). [CrossRef]

[7] Fan, D. P., Ji, G. P., Sun, G., Cheng, M. M., Shen, J., & Shao, L. (2020). Camouflaged object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2777-2787). [CrossRef]

[8] Usman, M. T., Khan, H., Rida, I., & Koo, J. (2025). Lightweight transformer-driven multi-scale trapezoidal attention network for saliency detection. *Engineering Applications of Artificial Intelligence, 155*, 110917. [CrossRef]

[9] Pei, J., Cheng, T., Fan, D. P., Tang, H., Chen, C., & Van Gool, L. (2022, October). Osformer: One-stage camouflaged instance segmentation with transformers. In *European conference on computer vision* (pp. 19-37). Cham: Springer Nature Switzerland. [CrossRef]

[10] Zheng, Y., Zhang, X., Wang, F., Cao, T., Sun, M., & Wang, X. (2018). Detection of people with camouflage pattern via dense deconvolution network. *IEEE Signal Processing Letters, 26*(1), 29-33. [CrossRef]

[11] Le, T. N., Nguyen, T. V., Nie, Z., Tran, M. T., & Sugimoto, A. (2019). Anabranch network for camouflaged object segmentation. *Computer vision and image understanding, 184*, 45-56. [CrossRef]

[12] Yang, Y., & Zhang, Q. (2023). Finding camouflaged objects along the camouflage mechanisms. *IEEE Transactions on Circuits and Systems for Video Technology, 34*(4), 2346-2360. [CrossRef]

[13] Song, Z., Kang, X., Wei, X., Liu, H., Dian, R., & Li, S. (2023). FSNet: Focus scanning network for camouflaged object detection. *IEEE Transactions on Image Processing, 32*, 2267-2278. [CrossRef]

[14] Sun, Y., Chen, G., Zhou, T., Zhang, Y., & Liu, N. (2021). Context-aware Cross-level Fusion Network for Camouflaged Object Detection. In *30th International Joint Conference on Artificial Intelligence, IJCAI 2021* (pp. 1025-1031). International Joint Conferences on Artificial Intelligence. [CrossRef]

[15] Hu, X., Wang, S., Qin, X., Dai, H., Ren, W., Luo, D., ... & Shao, L. (2023, June). High-resolution iterative feedback network for camouflaged object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 1, pp. 881-889). [CrossRef]

[16] Huang, Z., Dai, H., Xiang, T. Z., Wang, S., Chen, H. X., Qin, J., & Xiong, H. (2023, June). Feature Shrinkage Pyramid for Camouflaged Object Detection with Transformers. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5557-5566). IEEE. [CrossRef]

[17] Zhou, T., Zhou, Y., Gong, C., Yang, J., & Zhang, Y. (2022). Feature aggregation and propagation network for camouflaged object detection. *IEEE Transactions on Image Processing, 31*, 7036-7047. [CrossRef]

[18] Lyu, Y., Zhang, H., Li, Y., Liu, H., Yang, Y., & Yuan, D. (2023). UEDG: Uncertainty-edge dual guided camouflage object detection. *IEEE Transactions on Multimedia, 26*, 4050-4060. [CrossRef]

[19] Li, A., Zhang, J., Lv, Y., Liu, B., Zhang, T., & Dai, Y. (2021, June). Uncertainty-aware Joint Salient Object and Camouflaged Object Detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10066-10076). IEEE. [CrossRef]

[20] Zhou, X., Wu, Z., & Cong, R. (2024). Decoupling and integration network for camouflaged object detection. *IEEE Transactions on Multimedia, 26*, 7114-7129. [CrossRef]

[21] Song, Z., Kang, X., Wei, X., Liu, J., Lin, Z., & Li, S. (2025). Continuous feature representation for camouflaged object detection. *IEEE Transactions on Image Processing, 34*, 5672-5686. [CrossRef]

[22] Ji, H., Xie, F., Pan, L., Zheng, Y., & Shi, Z. (2025). HUNTNet: Homomorphic Unified Nexus Topology for Camouflaged Object Detection. *IEEE Transactions on Image Processing, 34*, 6068-6083. [CrossRef]

[23] Fan, D. P., Zhai, Y., Borji, A., Yang, J., & Shao, L. (2020, August). BBS-Net: RGB-D salient object detection with a bifurcated backbone strategy network. In *European conference on computer vision* (pp. 275-292). Cham: Springer International Publishing. [CrossRef]

[24] Khan, H., Usman, M. T., & Koo, J. (2025). Bilateral feature fusion with hexagonal attention for robust saliency detection under uncertain environments. *Information Fusion, 121*, 103165. [CrossRef]

[25] Yang, F., Zhai, Q., Li, X., Huang, R., Luo, A., Cheng, H., & Fan, D. P. (2021, October). Uncertainty-Guided Transformer Reasoning for Camouflaged Object Detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 4126-4135). IEEE. [CrossRef]

[26] Jia, Q., Yao, S., Liu, Y., Fan, X., Liu, R., & Luo, Z. (2022, June). Segment, Magnify and Reiterate: Detecting Camouflaged Objects the Hard Way. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4703-4712). IEEE. [CrossRef]

[27] Liu, Z., Zhang, Z., Tan, Y., & Wu, W. (2022, August). Boosting camouflaged object detection with dual-task interactive transformer. In *2022 26th International Conference on Pattern Recognition (ICPR)* (pp. 140-146). IEEE. [CrossRef]

[28] Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., & Fan, D. P. (2021, June). Simultaneously Localize, Segment and Rank the Camouflaged Objects. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11586-11596). IEEE. [CrossRef]

[29] Le, T. N., Nguyen, V., Le, C., Nguyen, T. C., Tran, M. T., & Nguyen, T. V. (2021, May). Camoufinder: Finding camouflaged instances in images. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 18, pp. 16071-16074). [CrossRef]

[30] Chen, Z., Sun, K., & Lin, X. (2024, March). CamoDiffusion: Camouflaged object detection via conditional diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 2, pp. 1272-1280). [CrossRef]

[31] He, C., Li, K., Zhang, Y., Xu, G., Tang, L., Zhang, Y., ... & Li, X. (2023). Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. *Advances in Neural Information Processing Systems, 36*, 30726-30737. [CrossRef]

Shadab Khan received BS degree in Computer Science from Islamia College Peshawar, Peshawar, Pakistan. He is currently pursuing a master's degree with the Department of Software Convergence, Sejong University, Seoul, South Korea. He is also working as a Research Assistant at Sejong University. His primary research interests include object detection and segmentation.

Abdurrahman Khan is a Computer Science student with excellent skills in programming, data structures, software development, and a growing expertise in machine learning, deep learning, and computer vision. Alongside his technical learning, he is also exploring graphic design to strengthen his creativity and visual communication skills. He is passionate about integrating technology and design to create innovative digital solutions, and his current focus is on building a strong profile that connects intelligent systems with user-friendly and visually engaging experiences.

Danish Ali is a Ph.D. student in Electrical and Computer Engineering at Villanova University, conducting research on advanced radar systems, artificial intelligence, and secure wireless communication technologies. His work integrates signal processing, microwave engineering, and deep learning to develop next-generation radar architectures for healthcare monitoring, autonomous vehicles, and intelligent security systems. With academic experience spanning Pakistan, China, and the United States, he has built a strong foundation in wireless communications, machine learning, embedded systems, and hardware design. His technical expertise includes mmWave Studio, GNU Radio, MATLAB, STM32, Arduino, and Verilog. Danish is passionate about advancing technologies at the intersection of AI, radar, and wireless systems to enhance global connectivity, safety, and sustainability.