PASR: Pattern-Aware Scene-Conditioned Reasoning for Camouflaged Object Detection

Xinyu Wang[1], Jintang Xue[1], and C.-C. Jay Kuo[1]*

University of Southern California, Los Angeles, CA 90089, USA
{xwang350,jintangx,jckuo}@usc.edu

Abstract. Camouflaged Object Detection (COD) remains challenging due to the deliberate alignment of foreground statistics with surrounding background patterns, which induces strong scene-dependent ambiguity. Existing approaches typically adopt object-centric modeling or reference-based augmentation. They often treat object appearance and background statistics independently, without explicitly modeling their conditional relationship. This work treats COD as a scene-conditioned pattern-deviation reasoning problem and captures how a camouflaged object deviates from its background scene. In this direction, we propose a pattern-aware scene-conditioned reasoning (PASR) method with two stages. The first stage constructs a scene-aligned background anchor via background-first retrieval, while the second stage performs a conditioned patch-level deviation reasoning. A reference prototype library is constructed in an offline and annotation-free manner, where foreground prototypes are derived from self-inferred rough masks rather than manually curated annotations. Unlike retraining-based paradigms, PASR achieves generalization through prototype-level expansion rather than parameter adaptation, enabling scalable inference across diverse background distributions. The resulting deviation maps can serve as priors for generic segmentation models. Extensive experiments on challenging COD benchmarks demonstrate that PASR outperforms existing unsupervised and weakly supervised methods and significantly narrows the gap with fully supervised models.

Keywords: Camouflaged Object Detection · Scene-Conditioned Reasoning · Prototype Learning

1 Introduction

Camouflaged Object Detection (COD) aims to identify intentionally concealed objects in scenes where the foreground and background share highly similar visual statistics [10]. Unlike conventional segmentation problems that rely on separable foreground cues, camouflage arises from deliberate mimicry in which color, texture, and structure are adjusted to align the object with its surroundings.

* Corresponding author.

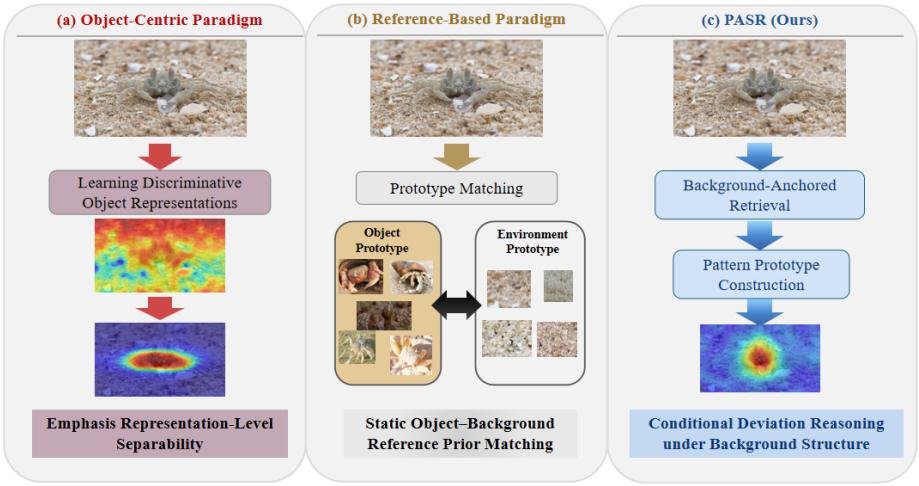


Fig. 1: Comparison of object-centric, reference-based, and PASR paradigms for COD. PASR shifts from foreground discrimination to modeling structured perturbations under scene-specific background statistics.

This mechanism induces strong scene-dependent ambiguity and substantial distributional overlap between foreground and background. Consequently, the core challenge of COD is not simply recognizing object appearance but reasoning about how an object deviates from its specific background context.

Existing approaches predominantly follow an object-centric paradigm, learning powerful object priors through diverse designs such as multi-scale fusion [3, 13, 45, 53] and boundary-aware design [25, 51, 54]. These methods implicitly assume that sufficiently expressive representations can separate foreground from background. Such learned object priors may require task-specific optimization to adapt to new background distributions. More recently, reference-based strategies [5, 11, 24] incorporate contextual cues through object references or background prototypes, but typically treat object appearance and background statistics independently, without explicitly modeling their conditional relationship.

These limitations motivate a mechanism-driven reformulation. In camouflaged scenes, foreground regions appear as structured perturbations over background statistics rather than as independent visual entities. In this view, the modeling objective shifts from learning standalone objects or background prototypes to capturing how deviation should manifest under a specific background structure. As illustrated in Fig. 1, we therefore reformulate COD as a scene-conditioned pattern-deviation reasoning problem: given a scene-level background condition, object hypotheses should be evaluated through coupled reasoning between background structures and object patterns.

In this direction, we propose PASR (Pattern-Aware Scene-conditioned Reasoning), which is a structured retrieval-based reasoning framework operating entirely in a frozen pretrained embedding space. In the first stage, a background-

first retrieval strategy identifies background-consistent references to construct a scene-specific background anchor (BG-anchor), establishing the contextual condition for subsequent inference. In the second stage, pattern-aware prototypes are constructed and matched under the BG-anchor, enabling detection through conditional deviation evidence rather than absolute appearance similarity.

Importantly, we observe that precise pixel-level supervision is not essential for effective deviation modeling. Object prototypes can be reliably derived from self-inferred coarse masks when conditioned on background structure. Accordingly, we construct a scene-conditioned reference library in an annotation-free manner, eliminating the need for manual mask curation or dataset-specific adaptation. Rather than relying on end-to-end retraining, PASR achieves generalization through prototype-level expansion and structured retrieval, enabling scalable adaptation to diverse background distributions. The resulting deviation maps can further serve as priors for geometric refinement in generic segmentation models, while the core detection signal remains grounded in background-conditioned reasoning.

Our main contributions are summarized as follows.

1. We reformulate COD as background-conditioned pattern-deviation reasoning, shifting from foreground discrimination to modeling structured perturbations under scene-specific background statistics.
2. We propose a Pattern-Aware Scene-conditioned Reasoning (PASR) framework that models the adaptive camouflage patterns through conditional background-object prototype deviations within a scene-aligned frozen embedding space.
3. We introduce an annotation-free reference construction strategy based on self-inferred coarse masks, enabling automatic, scalable construction of a prototype library without pixel-level supervision.

2 Related Work

2.1 Camouflaged Object Detection

Camouflaged Object Detection (COD) has attracted increasing attention due to the inherent challenge posed by the high visual similarity between the foreground and background [10, 22]. Most existing approaches follow an object-centric, discriminative paradigm, enhancing foreground separability through architectural refinement and feature interaction. Representative fully supervised methods improve representation from different perspectives, including frequency-aware modeling (*e.g.* FEDER [12]), progressive refinement and hierarchical feature learning (*e.g.* SINet [10], PFNet [29]), multi-scale integration (*e.g.* ZoomNet [32], ZoomNeXt [33]), graph-based interaction (*e.g.* HGINet [48]), and edge-aware coherence modeling (*e.g.* ICEG [15]). Despite diverse designs, such reliance on large-scale pixel-level supervision may limit generalization when background distributions shift or object categories vary. Recent unsupervised COD methods (*e.g.* SdalsNet [36], UCOD-DPL [47]) reduce the need for manual annotations

through self-distilled attention shifting or dynamic pseudo-label learning, but still optimize object-centric models on unlabeled COD training images. Overall, these object-centric methods implicitly assume that foreground and background can be separated through learned object representations. However, in camouflage scenarios, this assumption often breaks down.

To address the limitations of discriminative modeling, recent work introduces external references to assist in detection. Reference-driven COD methods can be broadly divided into object-semantic references and environment-aware modeling. The former leverages same-class exemplars to guide feature modulation or enrichment (*e.g.* R2CNet [52], UAT [46]), while the latter constructs background reference libraries to identify targets via deviation analysis (*e.g.* EASE [5]). Although these methods incorporate external priors, they typically model the object and background independently, without explicitly capturing their conditional coupling. Camouflage fundamentally arises from structural adaptation to the surrounding background, which makes detection inherently a conditional inference problem. This motivates our scene-conditioned deviation reasoning formulation, in which the conditional relationship between scene structure and object manifestation is explicitly modeled.

2.2 Prototype Learning and Retrieval-Based Visual Reasoning

Prototype learning formulates recognition as similarity matching to compact semantic representatives [39]. This paradigm has been widely adopted in dense prediction, particularly in a few-shot segmentation. Methods such as PANet [41] and PFENet [40] transfer foreground prototypes via metric alignment or prior-guided enrichment, while HSNet [30] and HDMNet [34] extend matching to hierarchical dense correspondence modeling. Retrieval-based frameworks further introduce non-parametric external references at inference time to enhance robustness. Methods such as RAC [26] and RALF [20] incorporate nearest-neighbor evidence to improve recognition in long-tailed or open-set settings. Large-scale pretrained models, including DINO [2], DINO-v2 [31], and SAM [21, 35], further amplify these paradigms by providing transferable representations and prompt-masking capabilities. In-context segmentation frameworks similarly condition predictions on reference examples without parameter updates. Despite these variations, a shared assumption persists: object identity can be summarized by relatively stable category-level evidence transferable across scenes. Whether represented as prototypes, retrieved instances, or prompts, references are primarily interpreted as auxiliary semantic signals that stabilize recognition.

Such semantic stability assumptions become problematic in camouflage scenarios. Camouflaged objects are not characterized by a globally distinctive appearance but by a structured deviation conditioned on specific background statistics. In this case, object representation is inherently scene-dependent rather than category-stable. Static prototype matching, even when augmented with retrieval or foundation features, cannot explicitly model this conditional structural dependency. We instead reformulate reference-based reasoning as background-anchored conditional deviation modeling, constructing scene-conditioned pro-

totypes that explicitly encode structured perturbations relative to background anchors.

3 Method

3.1 Rethinking Camouflage as Background-Imitative Pattern Formation

Camouflage arises fundamentally from imitation. Objects in camouflaged scenes adapt their appearance to mimic the surrounding background statistics, leading to strong overlap between the foreground and background feature distributions. This makes global foreground-background discrimination unreliable. More importantly, imitation is inherently contextual. When objects hide within a specific environment, their visual patterns are constrained by the same background statistics. Consequently, camouflaged objects under similar background conditions exhibit consistent structural tendencies.

This observation challenges the implicit assumption that the foreground-background discrepancy can be learned globally across scenes. In camouflage scenarios, deviation is background-dependent rather than universal. Instead of modeling foreground and background as independent entities, we reason about object evidence relative to aligned background structures. COD is thus formulated as background-conditioned pattern reasoning.

In this formulation, object hypotheses are evaluated within a context-aligned prototype space, where deviations are measured relative to matched background structures rather than a global foreground prior. Importantly, this reasoning does not rely on semantic supervision and operates in a fully unsupervised manner.

3.2 Overall Architecture

An overview of the proposed pattern-aware scene-conditioned reasoning (PASR) method is shown in Fig. 2. PASR provides a structured and annotation-free COD solution. Rather than direct separation in a global feature space, PASR performs inference within a progressively aligned reference subspace. It consists of two stages: (1) offline unsupervised prototype library construction and (2) progressive scene-conditioned reasoning at inference time. The prototype library encodes structured background-object configurations in a frozen embedding space. During inference, PASR first establishes a background-aligned reference and then performs pattern-conditioned deviation modeling followed by density-sensitive response separation. Through progressive alignment and associated pattern modeling, PASR formulates COD as structured deviation inference without manual supervision.

3.3 Unsupervised Prototype Library Construction

To establish a structured reference representation space for subsequent reasoning, we construct a foreground-background prototype library in an annotation-free and offline manner. Reference images are collected from publicly available

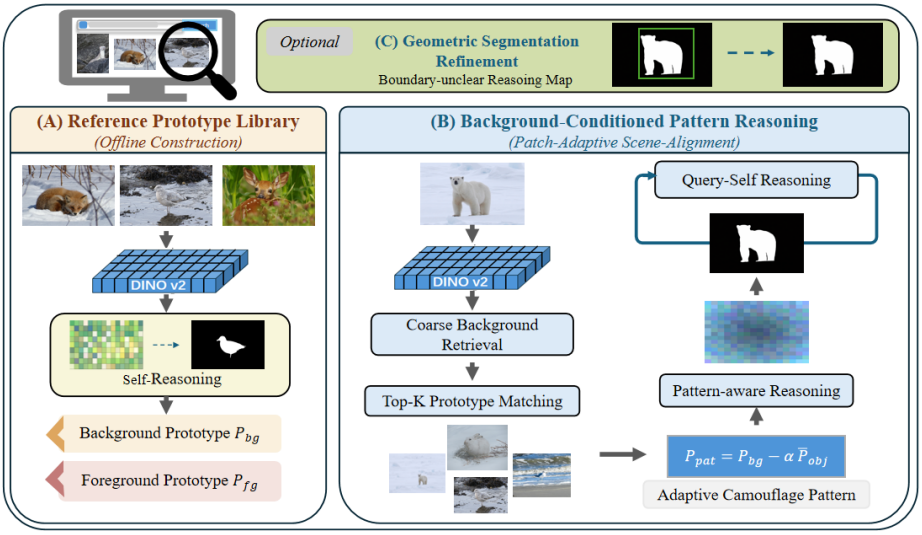


Fig. 2: An overview of PASR is depicted, which comprises an unsupervised prototype library and a progressive scene-conditioned reasoning module. The reference foreground-background prototype library is constructed offline. During inference, a background anchor and a coupled pattern refinement align a conditioned reference subspace progressively, in which object hypotheses are evaluated. An optional SAM-based refinement can be adopted to improve the segmentation.

natural scene platforms [19] to capture diverse background statistics. Only images released under Creative Commons licenses permitting research reuse are included, with all license-specific conditions respected. Unless otherwise specified, all main experiments rely on this external reference library without accessing any task-specific annotations.

Foreground-Background Decomposition via Self-Reasoning. For each reference image R_t, we extract dense embeddings $F_t \in \mathbb{R}^{H\times W\times d}$ using the frozen DINO-v2 backbone. We perform clustering in the embedding space, yielding a set of clusters $\mathcal{C}_t = \{C_t^k\}$. To obtain stable background statistics, clusters with large spatial extent and centroids located near the image center are excluded from background estimation. The remaining clusters are regarded as candidate background-consistent groups. For each candidate background cluster C_t^k, we define a local background prototype estimate $\tilde{P}_{t,k}^{bg}$ as the spatial expectation on the cluster. We then compute the cosine similarity between each spatial embedding and all candidate background prototypes as

$$S_t^k(x,y) = \mathrm{sim}\big(F_t(x,y), \tilde{P}_{t,k}^{bg}\big) = \frac{F_t(x,y)\cdot \tilde{P}_{t,k}^{bg}}{\|F_t(x,y)\|\|\tilde{P}_{t,k}^{bg}\|}. \tag{1}$$

The self-reasoning response at each location is defined as the maximum similarity over candidate background clusters:

$$S_t^{self}(x,y) = \max_k \; S_t^k(x,y). \tag{2}$$

Instead of relying on heuristic thresholds, we perform a distribution-aware separation by estimating the empirical density of the response map S_t^{self} through kernel density estimation (KDE). An adaptive threshold τ_t is determined from the dominant response mode, and the coarse foreground mask is obtained as

$$\tilde{M}_t(x,y) = \mathbf{1}\big(S_t^{self}(x,y) < \tau_t\big). \tag{3}$$

To enforce structural coherence and boundary precision, we further apply a geometric refinement stage based on SAM. The coarse mask $\tilde{M}_t$ and its derived bounding box are used as a prompt input to SAM. SAM operates with frozen parameters and is employed for structural regularization without introducing task-specific semantic supervision. The refined mask M_t is treated as a pseudo ground-truth signal for subsequent prototype extraction.

Foreground-Background Prototype Extraction. Given the pseudo ground-truth mask M_t, foreground and background prototypes are derived as masked spatial averages of pixel-level embeddings:

$$P_t^{fg} = \frac{\sum_x \sum_y M_t(x,\, y)\, F_t(x,\, y)}{\sum_x \sum_y M_t(x,\, y) + \epsilon}, \tag{4}$$

$$P_t^{bg} = \frac{\sum_x \sum_y (1 - M_t(x,\, y))\, F_t(x,\, y)}{\sum_x \sum_y (1 - M_t(x,\, y)) + \epsilon}. \tag{5}$$

Each reference image is represented by a prototype pair (P_t^{fg}, P_t^{bg}). The complete prototype library is constructed as

$$\mathcal{R} = \left\{ (P_t^{fg}, P_t^{bg}) \mid t = 1, \ldots, N_R \right\}. \tag{6}$$

The entire procedure, including clustering-based self-reasoning, distribution-aware separation, and geometric refinement, is conducted without manual annotations, producing an annotation-free prototype library that encodes structured background-object configurations for downstream reasoning.

3.4 Background-Conditioned Camouflage Pattern Reasoning

Camouflage is fundamentally a context-dependent phenomenon. The foreground is designed to mimic the surrounding environment, and thus object evidence should be interpreted relative to background statistics rather than as an absolute foreground prior. Accordingly, PASR decomposes reasoning into (i) coarse (image-level) background alignment to establish a scene-consistent reference, and (ii) fine-grained patch-level pattern matching.

Coarse Background Subspace Alignment (BG-anchor). Reasoning over the entire reference space may introduce cross-scene interference, since background statistics vary substantially across environments. We therefore first establish a scene-consistent reference subspace via coarse background alignment.

Given a query image I_j, we extract patch embeddings $\{F_j(x, y)\}$ using the same frozen DINO-v2 encoder and derive an image-wise global descriptor by spatial averaging:

$$F_j^g = \frac{1}{HW} \sum_x \sum_y F_j(x, y). \tag{7}$$

We then measure the similarity against all reference background prototypes:

$$S_{jt}^g = \mathrm{sim}\big(F_j^g, P_t^{bg}\big) = \frac{F_j^g \cdot P_t^{bg}}{\|F_j^g\| \|P_t^{bg}\|}. \tag{8}$$

and remove unmatched images, retaining the top-ranked candidates to form a subset $\mathcal{R}_j$. This subset functions as a scene-level background anchor (BG-anchor), constraining subsequent inference to structurally aligned reference configurations. This stage aims to enforce contextual consistency rather than localize the object.

Background-Conditioned Pattern Reasoning and Statistical Decomposition. Within the aligned subset $\mathcal{R}_j$, PASR performs a patch-level retrieval guided by background data followed by deviation-based reasoning. For each query patch embedding $F_j(x, y)$, we first retrieve the most-aligned reference image via its cosine similarity to the corresponding background prototypes:

$$S_{jt}^{bg}(x, y) = \mathrm{sim}\big(F_j(x, y),\, P_t^{bg}\big), \quad t \in \mathcal{R}_j. \tag{9}$$

For each spatial location (x, y), we select the references aligned to the top-K background $\mathcal{T}_j(x, y)$. Instead of directly matching with fixed pattern prototypes, we construct a query-adaptive object prototype by averaging the object embeddings of the selected references:

$$\bar{P}_j^{obj}(x, y) = \frac{1}{K} \sum_{t \in \mathcal{T}_j(x,y)} P_t^{fg}. \tag{10}$$

The background-conditioned pattern prototypes are constructed by combining the top-K background-conditioned pattern prototype and the query-adaptive object prototype:

$$P_{jt}^{pat}(x, y) = P_t^{bg} - \alpha\, \bar{P}_j^{obj}(x, y), \quad t \in \mathcal{T}_j(x, y) \tag{11}$$

where α controls the suppression strength of object-specific components. The deviation control parameter α is empirically selected based on the validation performance. In this formulation, camouflage cues are modeled as structured deviations from background distributions rather than independent foreground

signals. The resulting P_{jt}^{pat} serves as a reasoning primitive for evaluating deviation evidence under scene-conditioned constraints. The final deviation response at each patch location is computed as

$$S_j(x,y) = \frac{1}{K} \sum_{t \in \mathcal{T}_j(x,y)} \mathrm{sim}\big(F_j(x,y), P_{jt}^{pat}(x,y)\big). \tag{12}$$

This response encodes scene-conditioned camouflage evidence: lower similarity values indicate stronger agreement with background-conditioned deviation patterns, while higher responses correspond to background-consistent regions.

Then, we perform unsupervised statistical separation by estimating the response distribution of each image via kernel density estimation (KDE). An adaptive threshold is determined from the deviation-dominant region of the distribution, corresponding to low-similarity responses under the BG-anchor. This yields a reasoning-driven segmentation mask, $M_j^{(0)}$, highlighting regions that exhibit structured deviation from the background.

Query-Self Reasoning Refinement. The initial mask prediction may still be noisy due to imperfect cross-image prototype matching. We therefore perform an additional self-reasoning step using query-derived prototypes. Specifically, given the initial mask $M_j^{(0)}$, we compute query-specific background and foreground prototypes:

$$\hat{P}_j^{fg} = \frac{\sum_x \sum_y M_j^{(0)}(x,\,y)\, F_j(x,\,y)}{\sum_x \sum_y M_j^{(0)}(x,\,y) + \epsilon}, \quad \hat{P}_j^{bg} = \frac{\sum_x \sum_y (1 - M_j^{(0)}(x,\,y))\, F_j(x,\,y)}{\sum_x \sum_y (1 - M_j^{(0)}(x,\,y)) + \epsilon}, \tag{13}$$

We then classify each query patch by contrasting its cosine similarity to the two prototypes:

$$\Delta_j(x,y) = \mathrm{sim}\big(F_j(x,\,y), \hat{P}_j^{bg}\big) - \mathrm{sim}\big(F_j(x,\,y), \hat{P}_j^{fg}\big), \tag{14}$$

Finally, we apply the same KDE-based separation on Δ_j to obtain a refined mask $M_j^{(1)}$, which suppresses the noise of cross-scene search by enforcing the consistent statistical decomposition of queries.

In general, PASR integrates background anchoring, background-conditioned pattern primitives, and distribution-sensitive statistical separation into a unified reasoning framework. The entire process operates without global foreground priors or category-level supervision, enabling scene-conditioned camouflage inference in a fully unsupervised manner.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate PASR on four widely-used camouflaged object detection (COD) benchmarks: CHAMELEON [38], CAMO [22], COD10K [10] and

NC4K [27]. The CHAMELEON [38] dataset consists of 76 camouflaged images with pixel-wise annotations and is commonly used as a small-scale evaluation benchmark. The CAMO [22] dataset contains 1,250 camouflaged images, including 1,000 for training and 250 for testing. COD10K [10] is a large-scale COD benchmark that covers 78 object categories in diverse natural scenes, including 3,040 training images and 2,026 testing images. The NC4K [27] dataset comprises 4,121 natural camouflaged images with complex background interference and greater diversity of the scene. During evaluation, we report the results on the testing sets of CAMO [22] and COD10K [10], the entire CHAMELEON [38] dataset, and the complete NC4K [27] dataset following standard protocols [5,10].

Evaluation Metrics. We adopt four widely-used quantitative metrics for performance evaluation: S-measure (S_m) [7], mean absolute error (MAE), enhanced alignment measure (E_ϕ) [8], and weighted F-measure (F_β^w) [28]. Among them, higher values of S_m, E_ϕ, and F_β^w indicate better structural and alignment quality of the predicted masks, while lower MAE reflects better detection.

Implementation Details. PASR adopts a frozen DINOv2-ViT-L/14 [31] encoder to extract patch-level embeddings for both reference library construction and query inference. The same backbone is used consistently throughout all stages to ensure feature-space alignment between reference prototypes and test images. For reference library construction, foreground-background decomposition is achieved via the proposed unsupervised self-reasoning mechanism. This process is conducted entirely offline and does not rely on ground-truth annotations, dataset-specific tuning, or semantic supervision. To enhance the structural coherence of the coarse masks obtained during this stage, we employ SAM-ViT-B [21] as a geometric refinement module. Importantly, SAM is used purely for mask-level structural regularization without introducing semantic priors.

PASR adopts a progressive background-guided retrieval strategy. We set the number of reference prototypes of the best candidate retrieved for pattern reasoning to $K = 32$. The coarse BG-anchor is built by eliminating 50% of candidate references through global background retrieval. The control parameter α is set to 0.4. On a single NVIDIA A40, the core inference takes 71–76 ms per image: ∼51 ms for DINO-v2, ∼13 ms for reasoning, and ∼10 ms for query self-refinement. An optional boundary refinement stage further improves structural precision by using the reasoning-derived mask as a prompt for refinement.

4.2 Comparison with State-of-the-art Methods

Comparison Methods. To comprehensively evaluate PASR under different supervision paradigms, we report the results in two evaluation settings and compare them with the corresponding baselines. We first evaluate PASR in a pure reasoning setting, where segmentation is obtained through prototype-based structured reasoning without any SAM refinement. Benchmarking methods include training-free or unsupervised object detection methods, such as DiffCut [4], TokenCut [44], MaskCut [43], FOUND [37], VoteCut [1], FreeSOLO [42],

Table 1: Quantitative comparisons with representative unsupervised methods on COD datasets. For DINO-based methods, we adopt a unified DINOv2-ViT-L/14 backbone for fair comparison. The best and second-best results are in bold and with an underline, respectively.

Method	CHAMELEON				CAMO				COD10K				NC4K			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$
FreeSOLO [42]	0.659	0.730	0.503	0.101	0.534	0.600	0.333	0.177	0.630	0.725	0.421	0.093	0.611	0.781	0.441	0.128
DiffCut [4]	0.537	0.592	0.374	0.245	0.630	0.683	0.491	0.166	0.591	0.628	0.379	0.160	0.671	0.727	0.531	0.134
MaskCut [43]	0.687	0.719	0.531	0.115	0.630	0.658	0.494	0.194	0.574	0.609	0.353	0.190	0.646	0.668	0.503	0.177
TokenCut [44]	0.682	0.705	0.494	0.088	0.620	0.655	0.439	0.144	0.619	0.649	0.373	0.088	0.664	0.699	0.492	0.108
FOUND [37]	0.659	0.745	0.490	0.104	0.558	0.622	0.373	0.172	0.605	0.675	0.386	0.100	0.621	0.693	0.455	0.131
VoteCut [1]	0.657	0.677	0.474	0.112	0.526	0.522	0.303	0.196	0.619	0.646	0.389	0.103	0.609	0.623	0.412	0.140
RISE [6]	0.822	0.884	0.720	0.050	0.734	0.787	0.610	0.109	0.763	0.840	0.600	0.049	0.805	0.868	0.705	0.061
EASE [5]	0.819	0.899	0.741	0.044	0.749	0.831	0.684	**0.098**	0.773	0.866	0.656	**0.040**	0.800	0.884	0.735	0.056
Ours (PASR)	**0.832**	**0.924**	**0.768**	**0.041**	**0.755**	**0.833**	**0.693**	0.102	**0.774**	**0.868**	**0.664**	0.048	**0.817**	**0.902**	**0.760**	**0.051**

RISE [6] and EASE [5]. For a fair comparison among unsupervised baselines, we unify the underlying principle of DINO-based methods by adopting DINOv2-ViT-L/14 [31] whenever applicable. Following [5], all methods are evaluated under identical input resolution 476×476 and without test-time augmentation and post-processing techniques to ensure consistency.

We further evaluate PASR under the SAM-based refinement setting, where the reasoning-derived segmentation mask is converted into bounding box prompts for SAM. This setting allows us to assess the compatibility of PASR with prompt-driven segmentation paradigms. We compare against recent prompt-based methods, such as GenSAM [16], ProMaC [17], as well as RISE [6] and EASE [5] refined with SAM-based refinement. For reference, we also report results from representative weakly supervised methods, including SCWS [50], TEL [23], WS-SAM [13], and SEE [14], as well as fully supervised COD models such as SINet-v2 [9], ZoomNet [32], FEDER [12], CamoFormer [49], and HitNet [18]. All baseline results are obtained from official repositories or reproduced using publicly released implementations under the same evaluation protocol described in Section 4.1.

Quantitative Comparison. Compared to existing unsupervised methods (Table 1), PASR consistently achieves the best performance on all four COD datasets and outperforms previous state-of-the-art methods in nearly all evaluation metrics. In particular, compared to the two most recent unsupervised COD methods, RISE [6] and EASE [5], PASR demonstrates clear and consistent improvements. Performance gains are especially pronounced on CHAMELEON and NC4K, indicating stronger scene-level alignment and more accurate reasoning. These results validate the effectiveness of modeling camouflage as a background-conditioned pattern deviation, where object evidence is inferred relative to aligned background statistics rather than treated as an independent foreground prior.

As shown in Table 2, PASR further demonstrates strong competitiveness against weakly-supervised and prompt-based segmentation methods. Compared with weakly supervised baselines, PASR produces consistent improvements across all datasets, suggesting the strength of its structured background deviation mod-

Table 2: Quantitative comparisons across different supervision regimes, including fully-supervised, weakly-supervised, and prompt-based COD methods. SAM-ViT-H backbones are adopted for each SAM-based prompt method. The best and second-best results within each group are in bold and underlined, respectively.

Method	CHAMELEON				CAMO				COD10K				NC4K			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$
Fully-Supervised																
SINet-v2 [9]	0.888	0.942	0.816	0.030	0.820	0.882	0.743	0.070	0.815	0.887	0.680	0.037	0.847	0.914	0.805	0.048
ZoomNet [32]	0.902	0.958	0.845	0.023	0.820	0.892	0.752	0.066	0.838	0.911	0.729	0.029	0.853	0.912	0.784	0.043
FEDER [12]	0.887	0.954	0.834	0.030	0.802	0.867	0.738	0.071	0.822	0.900	0.716	0.032	0.847	0.907	0.789	0.044
HitNet [18]	**0.921**	**0.967**	**0.897**	**0.019**	0.849	0.906	0.809	0.055	**0.871**	**0.935**	**0.806**	**0.023**	0.875	0.926	0.834	0.037
CamoFormer [49]	0.909	0.956	0.871	0.022	**0.872**	**0.931**	**0.831**	**0.046**	0.869	0.931	0.786	**0.023**	**0.891**	**0.941**	**0.847**	**0.030**
Weakly-Supervised Methods																
SCWS [50]	0.792	0.881	0.717	0.053	0.713	0.658	0.648	0.102	0.710	0.805	0.634	0.055	0.784	0.814	0.705	0.073
TEL [23]	0.785	0.827	0.700	0.073	0.717	0.681	0.644	0.104	0.724	0.801	0.645	0.057	0.782	0.799	0.689	0.075
WS-SAM [13]	0.820	0.887	0.723	0.048	0.759	0.814	0.667	0.092	0.803	0.877	0.680	0.038	0.829	0.886	0.757	0.052
SEE [14]	**0.826**	**0.903**	**0.773**	**0.044**	**0.765**	**0.826**	**0.731**	**0.090**	**0.807**	**0.883**	**0.721**	**0.036**	**0.836**	**0.891**	**0.798**	**0.051**
Prompt-based Segmentation																
GenSAM [16]	0.767	0.827	0.673	0.075	0.738	0.803	0.674	0.106	0.773	0.832	0.667	0.065	0.810	0.866	0.751	0.065
ProMaC [17]	0.833	0.899	0.790	0.044	0.767	**0.846**	0.725	**0.090**	0.805	0.876	0.716	0.042	0.812	0.862	0.711	0.065
RISE [6]	0.823	0.882	0.733	0.055	0.760	0.807	0.651	0.102	0.790	0.854	0.643	0.044	0.825	0.874	0.736	0.056
EASE [5]	0.853	0.912	0.812	0.044	**0.777**	0.841	**0.748**	0.099	**0.856**	**0.914**	**0.801**	**0.028**	0.849	0.902	0.815	0.051
Ours (PASR)	**0.869**	**0.937**	**0.846**	**0.038**	0.768	0.840	0.734	0.105	0.853	0.913	**0.801**	0.034	**0.863**	**0.917**	**0.840**	**0.046**

Fig. 3: Examples from the constructed reference library. The rough masks obtained via self-reasoning are highlighted in green, while the bounding-box prompts used for geometric refinement are indicated in red.

eling. Moreover, although several prompt-based methods incorporate SAM refinement or additional guidance mechanisms, PASR relies solely on annotation-free prototype reasoning while achieving superior performance over recent prompt-driven methods such as ProMaC [17] and GenSAM [16]. It is worth noting that PASR also narrows the performance gap to fully supervised methods. On CHAMELEON, COD10K, and NC4K, PASR achieves performance comparable to fully supervised methods, and in several metrics even exceeds a number of supervised baselines. These results suggest that background-conditioned reasoning provides a strong inductive bias, significantly reducing the reliance on pixel-level annotations.

Qualitative Comparison. We visualize several coarse masks generated during reference library construction via self-reasoning and geometric refinement in Fig. 3. Although these rough masks do not capture fine-grained boundaries accurately, they provide reliable object locations and are structurally sufficient to derive meaningful prototypes and effectively guide subsequent scene-conditioned deviation reasoning. As shown in Fig. 4, we compare PASR with representative unsupervised segmentation methods. PASR achieves more accurate localization and cleaner object delineation, particularly in complex environments.

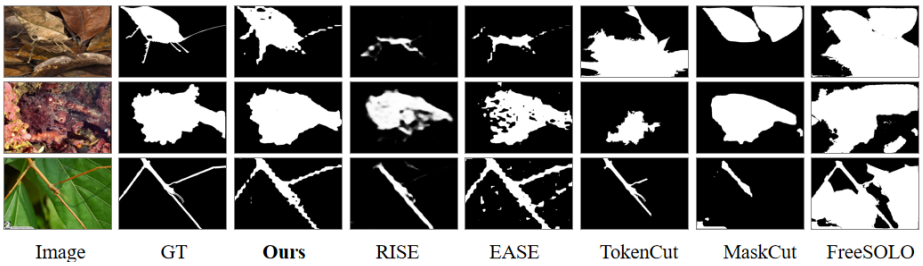
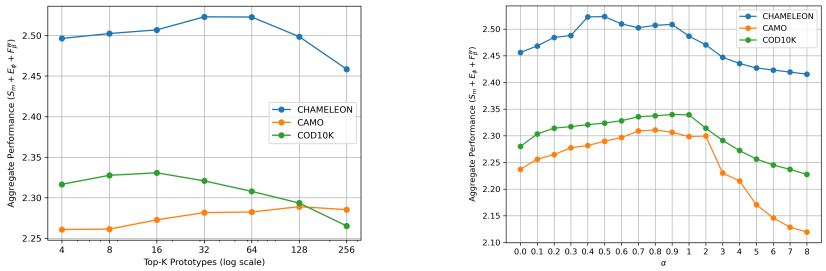


Fig. 4: Qualitative comparisons with the state-of-the-art unsupervised methods.



(a) Top-K sensitivity **(b)** Effect of α

Fig. 5: Parameter sensitivity analysis of PASR: (a) Sensitivity to the number of retrieved prototypes K, and (b) Effect of deviation coefficient α.

4.3 Ablation Study

Component Ablations. Table 3 evaluates the contribution of each component in PASR. When α=0, PASR degenerates to pure background matching and shows the largest overall drop, highlighting the importance of modeling foreground evidence as background-conditioned deviation. Patch-adaptive object prototypes and query self-refinement both improve the results, indicating that query-specific pattern evidence helps stabilize the predicted masks. In addition, replacing the BG-anchor with random background anchors results in clear degradation, showing that the improvement comes from scene-aligned background conditioning rather than merely a reduced patch-level search space. Notably, removing offline SAM during reference-library construction causes only minor changes, suggesting that SAM mainly refines the coarse prototype masks and is not the core source of the gain.

Deviation Hyperparameter Analysis. We further analyze the deviation coefficient α in Eq. (11). As α increases from 0, performance improves and remains stable within $\alpha \in [0.4, 1]$, forming a broad plateau across benchmarks (Fig. 5b). This suggests that explicitly modeling foreground components as structured deviations over background statistics is essential for camouflage reasoning, while the conditioned deviation formulation is robust rather than highly sensitive to

Table 3: Component ablations of PASR. Results are reported as S_α/F_β^w.

Variant	CHAM.	CAMO	COD10K	NC4K
$\alpha = 0$	0.819/0.745	0.741/0.674	0.767/0.654	0.798/0.734
w/o adapt. obj	0.825/0.755	0.753/0.691	0.773/0.662	0.813/0.755
w/o query-self	0.826/0.756	0.751/0.687	0.769/0.656	0.808/0.748
w/o BG-anchor	0.829/0.764	0.752/0.689	0.776/0.664	0.816/0.759
rand. BG-anchor	0.820/0.744	0.750/0.687	0.770/0.656	0.815/0.757
w/o offline SAM	0.829/0.764	0.754/0.692	0.774/0.662	0.812/0.752
Full PASR	0.832/0.768	0.755/0.693	0.774/0.664	0.817/0.760

Table 4: Comparison of reference libraries constructed from different image resources in terms of library scale, retrieval size, and prototype formulation. Notably, PASR achieves strong performance with a significantly smaller library and retrieval size.

Image Resource	**Library Size**	**Retrieval Size**	**Method**	**Prototype Formulation**
COD Dataset [10, 22]	4,040	32	PASR	Background-conditioned pattern prototype
Web Images	315	32	PASR	Background-conditioned pattern prototype
COD Dataset [10, 22]	4,040	512	RISE	Foreground & background prototypes
DiffPro [5]	16,000	1,024	EASE	Environmental prototype

precise coefficient tuning. However, overly large α values gradually degrade performance, likely due to excessive suppression that disrupts background-object structural coupling. Unless otherwise specified, we set $\alpha = 0.4$ across all datasets.

Reference Efficiency and Scalability. We evaluate the influence of retrieval size K and reference construction. As shown in Fig. 5a, PASR achieves near-optimal performance with a small retrieval size ($K = 32$) and remains stable over a wide range of K. Table 4 compares different reference-library settings. PASR retrieves top-32 candidates from only 315 web reference images, yet achieves competitive performance compared with prior unsupervised methods [5, 6] that rely on thousands of dataset-specific references and large retrieval budgets. Table 5 further shows that web-based references achieve comparable performance to COD-training references, and that limited point annotations bring only marginal gains. Together, these results suggest that PASR generalizes through structured background-conditioned prototype reasoning rather than dataset-specific supervision or large-scale reference accumulation.

5 Conclusion

This work addressed the COD problem from a background-deviation perspective. Instead of modeling camouflage as foreground saliency, we formulated it as a structured deviation from background statistics and proposed the PASR method. PASR outperformed existing unsupervised and weakly-supervised methods, and narrowed the performance gap to fully supervised methods. These results were achieved without pixel-level supervision, dataset-specific adaptation, or prompt-dependent segmentation. Our findings suggest that structured scene-conditioned

Table 5: PASR performance with reference libraries built from COD datasets or web images ($\mathcal{R}_1/\mathcal{R}_2$); * includes three-point annotations during prototype construction.

Method	CHAMELEON				CAMO				COD10K			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$
PASR-$\mathcal{R}_1^*$	**0.832**	0.919	**0.768**	**0.041**	**0.765**	**0.840**	**0.707**	**0.094**	**0.787**	**0.885**	**0.682**	**0.038**
PASR-$\mathcal{R}_1$	0.816	0.893	0.742	0.043	0.749	0.827	0.681	0.100	0.778	0.877	0.669	0.042
PASR-$\mathcal{R}_2^*$	0.829	0.913	0.760	0.042	0.755	0.836	0.694	0.101	0.784	0.881	0.677	0.042
PASR-$\mathcal{R}_2$	**0.832**	**0.924**	**0.768**	**0.041**	0.755	0.833	0.693	0.102	0.774	0.868	0.664	0.048

reasoning provides a strong inductive bias for camouflage modeling, reducing the dependency on large-scale annotations. This viewpoint encourages further exploration of context-aligned reasoning as a scalable alternative to supervision-driven paradigms in challenging visual recognition tasks.

References

1. Arica, S., Rubin, O., Gershov, S., Laufer, S.: CuVLER: Enhanced unsupervised object discoveries through exhaustive self-supervised transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23105–23114 (2024)
2. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9650–9660 (2021)
3. Chen, G., Liu, S.J., Sun, Y.J., Ji, G.P., Wu, Y.F., Zhou, T.: Camouflaged object detection via context-aware cross-level fusion. IEEE Transactions on Circuits and Systems for Video Technology **32**(10), 6981–6993 (2022)
4. Couairon, P., Shukor, M., Haugeard, J.E., Cord, M., Thome, N.: Diffcut: Catalyzing zero-shot semantic segmentation with diffusion features and recursive normalized cut. Advances in Neural Information Processing Systems **37**, 13548–13578 (2024)
5. Du, J., Hao, F., Yu, M., Kong, D., Wu, J., Wang, B., Xu, J., Li, P.: Shift the lens: Environment-aware unsupervised camouflaged object detection. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 19271–19282 (2025)
6. Du, J., Wang, X., Hao, F., Yu, M., Chen, C., Wu, J., Wang, B., Xu, J., Li, P.: Beyond single images: Retrieval self-augmented unsupervised camouflaged object detection. In: IEEE International Conference on Computer Vision. pp. 22131–22142 (2025)
7. Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: Proceedings of the IEEE international conference on computer vision. pp. 4548–4557 (2017)
8. Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M., Borji, A.: Enhanced-alignment measure for binary foreground map evaluation. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. pp. 698–704. IJCAI'18, AAAI Press (2018)
9. Fan, D.P., Ji, G.P., Cheng, M.M., Shao, L.: Concealed object detection. IEEE transactions on pattern analysis and machine intelligence **44**(10), 6024–6042 (2022)
10. Fan, D.P., Ji, G.P., Sun, G., Cheng, M.M., Shen, J., Shao, L.: Camouflaged object detection. In: IEEE CVPR. pp. 2777–2787 (2020)

11. Gupta, A., Jerripothula, K.R., Tillo, T.: CIRCOD: co-saliency inspired referring camouflaged object discovery. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 8313–8323. IEEE (2025)
12. He, C., Li, K., Zhang, Y., Tang, L., Zhang, Y., Guo, Z., Li, X.: Camouflaged object detection with feature decomposition and edge reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 22046–22055 (2023)
13. He, C., Li, K., Zhang, Y., Xu, G., Tang, L., Zhang, Y., Guo, Z., Li, X.: Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. Advances in Neural Information Processing Systems **36**, 30726–30737 (2023)
14. He, C., Li, K., Zhang, Y., Yang, Z., Pang, Y., Tang, L., Fang, C., Zhang, Y., Kong, L., Li, X., Farsiu, S.: Segment concealed objects with incomplete supervision. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(9), 7832–7851 (2025)
15. He, C., Li, K., Zhang, Y., Zhang, Y., You, C., Guo, Z., Li, X., Danelljan, M., Yu, F.: Strategic preys make acute predators: Enhancing camouflaged object detectors by generating camouflaged objects. In: International Conference on Learning Representations. pp. 36657–36675 (2024)
16. Hu, J., Lin, J., Gong, S., Cai, W.: Relax image-specific prompt requirement in sam: A single generic prompt for segmenting camouflaged objects. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 12511–12518 (2024)
17. Hu, J., Lin, J., Yan, J., Gong, S.: Leveraging hallucinations to reduce manual prompt dependency in promptable segmentation. Advances in Neural Information Processing Systems **37**, 107171–107197 (2024)
18. Hu, X., Wang, S., Qin, X., Dai, H., Ren, W., Luo, D., Tai, Y., Shao, L.: High-resolution iterative feedback network for camouflaged object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 881–889 (2023)
19. iNaturalist: inaturalist. Available from `https://www.inaturalist.org` (2026), accessed 27 February 2026
20. Kim, J., Cho, E., Kim, S., Kim, H.J.: Retrieval-augmented open-vocabulary object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17427–17436 (2024)
21. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
22. Le, T.N., Nguyen, T.V., Nie, Z., Tran, M.T., Sugimoto, A.: Anabranch network for camouflaged object segmentation. Computer vision and image understanding **184**, 45–56 (2019)
23. Liang, Z., Wang, T., Zhang, X., Sun, J., Shen, J.: Tree energy loss: Towards sparsely annotated semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16907–16916 (2022)
24. Liu, X., Huang, S., Wu, R., Zhao, H., Xu, D., Wei, X., Han, J., Liu, S.: Reference prompted model adaptation for referring camouflaged object detection. In: 2024 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2024)
25. Liu, Z., Jiang, P., Lin, L., Deng, X.: Edge attention learning for efficient camouflaged object detection. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5230–5234. IEEE (2024)

26. Long, A., Yin, W., Ajanthan, T., Nguyen, V., Purkait, P., Garg, R., Blair, A., Shen, C., Van den Hengel, A.: Retrieval augmented classification for long-tail visual recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6959–6969 (2022)
27. Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., Fan, D.P.: Simultaneously localize, segment and rank the camouflaged objects. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11591–11601 (2021)
28. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 248–255 (2014)
29. Mei, H., Ji, G.P., Wei, Z., Yang, X., Wei, X., Fan, D.P.: Camouflaged object segmentation with distraction mining. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8772–8781 (2021)
30. Min, J., Kang, D., Cho, M.: Hypercorrelation squeeze for few-shot segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6941–6952 (2021)
31. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: DINOv2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)
32. Pang, Y., Zhao, X., Xiang, T.Z., Zhang, L., Lu, H.: 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. pp. 2160–2170 (2022)
33. Pang, Y., Zhao, X., Xiang, T.Z., Zhang, L., Lu, H.: ZoomNeXt: A unified collaborative pyramid network for camouflaged object detection. IEEE transactions on pattern analysis and machine intelligence **46**(12), 9205–9220 (2024)
34. Peng, B., Tian, Z., Wu, X., Wang, C., Liu, S., Su, J., Jia, J.: Hierarchical dense correlation distillation for few-shot segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 23641–23651 (2023)
35. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollar, P., Feichtenhofer, C.: Sam 2: Segment anything in images and videos. In: Yue, Y., Garg, A., Peng, N., Sha, F., Yu, R. (eds.) International Conference on Learning Representations. vol. 2025, pp. 28085–28128 (2025)
36. Shou, P., Liu, Y., Wang, W., Sun, Y., Zheng, Z., Zhu, S., Yan, C.: Sdalsnet: self-distilled attention localization and shift network for unsupervised camouflaged object detection. In: AAAI. vol. 39, pp. 6914–6921 (2025)
37. Siméoni, O., Sekkat, C., Puy, G., Vobecký, A., Zablocki, E., Pérez, P.: Unsupervised object localization: Observing the background to discover objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3176–3186 (June 2023)
38. Skurowski, P., Abdulameer, H., Błaszczyk, J., Depta, T., Kornacki, A., Kozieł, P.: Animal camouflage analysis: Chameleon database. Unpublished manuscript **2**(6), 7 (2018)
39. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in neural information processing systems **30** (2017)
40. Tian, Z., Zhao, H., Shu, M., Yang, Z., Li, R., Jia, J.: Prior guided feature enrichment network for few-shot segmentation. IEEE transactions on pattern analysis and machine intelligence **44**(2), 1050–1065 (2022)

41. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: Panet: Few-shot image semantic segmentation with prototype alignment. In: proceedings of the IEEE/CVF international conference on computer vision. pp. 9197–9206 (2019)
42. Wang, X., Yu, Z., De Mello, S., Kautz, J., Anandkumar, A., Shen, C., Alvarez, J.M.: Freesolo: Learning to segment objects without annotations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14176–14186 (2022)
43. Wang, X., Girdhar, R., Yu, S.X., Misra, I.: Cut and learn for unsupervised object detection and instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3124–3134 (2023)
44. Wang, Y., Shen, X., Hu, S.X., Yuan, Y., Crowley, J.L., Vaufreydaz, D.: Self-supervised transformers for unsupervised object discovery using normalized cut. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14543–14553 (2022)
45. Wang, Y., Bi, X., Liu, B., Wei, Y., Li, W., Xiao, B.: Learning discriminative representations from cross-scale features for camouflaged object detection. IEEE Transactions on Circuits and Systems for Video Technology **34**(12), 12756–12769 (2024)
46. Wu, R., Xiang, T.Z., Xie, G.S., Gao, R., Shu, X., Zhao, F., Shao, L.: Uncertainty-aware transformer for referring camouflaged object detection. IEEE Transactions on Image Processing **34**, 5341–5354 (2025)
47. Yan, W., Chen, L., Kou, H., Zhang, S., Zhang, Y., Cao, L.: UCOD-DPL: Unsupervised camouflaged object detection via dynamic pseudo-label learning. In: CVPR. pp. 30365–30375 (2025)
48. Yao, S., Sun, H., Xiang, T.Z., Wang, X., Cao, X.: Hierarchical graph interaction transformer with dynamic token clustering for camouflaged object detection. IEEE Transactions on Image Processing **33**, 5936–5948 (2024)
49. Yin, B., Zhang, X., Fan, D.P., Jiao, S., Cheng, M.M., Van Gool, L., Hou, Q.: Camoformer: Masked separable attention for camouflaged object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), 10362–10374 (2024)
50. Yu, S., Zhang, B., Xiao, J., Lim, E.G.: Structure-consistent weakly supervised salient object detection with local saliency coherence. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 3234–3242 (2021)
51. Yue, G., Wu, S., Zhou, T., Li, G., Du, J., Luo, Y., Jiang, Q.: Progressive region-to-boundary exploration network for camouflaged object detection. IEEE Transactions on Multimedia **27**, 236–248 (2025)
52. Zhang, X., Yin, B., Lin, Z., Hou, Q., Fan, D.P., Cheng, M.M.: Referring camouflaged object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(5), 3597–3610 (2025)
53. Zheng, D., Zheng, X., Yang, L.T., Gao, Y., Zhu, C., Ruan, Y.: Mffn: Multi-view feature fusion network for camouflaged object detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 6232–6242 (2023)
54. Zhu, H., Li, P., Xie, H., Yan, X., Liang, D., Chen, D., Wei, M., Qin, J.: I can find you! boundary-guided separated attention network for camouflaged object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 3608–3616 (2022)