Journal of Algorithms and Computations

journal homepage: http://jac.ut.ac.ir

Camouflaged Object Segmentation: A Systematic Review of Methods and Evaluation

A. Jelokhani[†]

[1] School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

ABSTRACT

Camouflaged Object Segmentation (COS) aims to segment objects at the pixel level where objects have minimal visual contrast with their surroundings. Object boundaries are often weak, and background distractors can cause incomplete masks and false positives. Though deep learning techniques have significantly boosted the development of COS, literature is scattered and heterogeneous with different deep learning architectures, training procedures, and benchmark protocols. To address this, we conducted this systematic literature review to synthesize existing literature on Camouflaged Object Segmentation techniques published between 2020 and 2026. We analyzed 38 eligible studies, extracted standardized study characteristics (datasets/protocols, core methodological ideas, architectural choices, and reported improvements), and summarized them in a method-centric taxonomy and an appraisal table. The literature shows that consistent improvements are achieved through boundary-aware and structure-aware refinement, multi-scale/coarse-to-fine reasoning, transformer-based global context and local detail recovery, and uncertainty-aware iterative refinement.

Keywords: camouflaged object segmentation, PRISMA, segmentation, multi-scale learning, transformers, diffusion models.

AMS subject classification: 90C06

ARTICLE INFO

Article history:
Review Article
Received 16, April 2026
Accepted 10, June 2026
Available online 17, June 2026

[†] Corresponding author: A. Jelokhani, Email: ali.jelokhani@ut.ac.ir

1 Introduction

The task of camouflaged object segmentation (COS) is to separate objects with appearance showing very minor contrast with the surrounding background. Unlike other segmentation tasks, the task of camouflaged object segmentation is usually confronted with challenges such as weak or ambiguous boundaries, strong texture similarity, and background distractions. Therefore, the task of camouflaged object segmentation cannot be resolved using appearance features alone; instead, stronger inductive biases for boundary, context, and long-range dependencies must be considered.

To address the challenges of the task of camouflaged object segmentation, many deep learning strategies have been proposed in the literature for the task of camouflaged object segmentation. Some of the strategies proposed in the literature for the task of camouflaged object segmentation include boundary-guided refinement, edge-aware learning strategies [1, 2], multi-scale or coarse-to-fine reasoning for the task of camouflaged object segmentation [3, 4], CNN-Transformer cooperation for the task of camouflaged object segmentation [5], and the use of the diffusion or the unfolding framework for the task of camouflaged object segmentation [6-8]. The literature for the task of camouflaged object segmentation is difficult to consolidate because of the large variety of strategies proposed for the task of camouflaged object segmentation in the literature.

Consequently, this systematic literature review aggregates research on COS from the years 2020-2026 with the aim of establishing trends within the methodologies, assessing common practices within experimental methodologies, and identifying gaps that need future research. The contributions of this review are as follows: (i) a method-centric taxonomy, where we group methods within COS using primary modeling mechanisms, (ii) a structured appraisal with respect to reporting, where we assess strengths and limitations with respect to reproducibility and comparability, and (iii) a synthesis with respect to common practices, where we aggregate common performance drivers and trade-offs from the literature. Based on our contributions, we aim to analyze the datasets and evaluation practices used, as well as their implications, and assess the strengths and limitations with respect to method families, highlighting gaps that need future research. The review process was carried out using a set of best practices that align with the PRISMA protocol.

2 Methods

i) Eligibility Criteria

In order for the selected evidence to address camouflaged object segmentation and allow a consistent synthesis, inclusion and exclusion criteria were set.

ii) Inclusion Criteria:

The inclusion criteria for studies were:

- **Topic Relevance:** The proposed, assessed, or optimized approach in this research work relates to camouflaged object segmentation, which involves pixel-level segmentation of camouflaged objects in a camouflage or camouflage-like context.
- **Task Focus:** Firstly, the major task focus involves segmentation and not mere or only detection/localization, like bounding boxes or saliency maps, which do not necessarily predict masks.
- **Terminology note (COS vs. COD):** In the literature, camouflaged object segmentation and camouflaged object detection have been used interchangeably in some cases. For the purpose of this review, the selection of papers was based on the formulation/evaluation of the task instead of the title of the papers. Papers that use the term COD were selected only when they

performed the task of mask prediction/segmentation; papers that performed detection, bounding box, localization, or classification without mask prediction/segmentation were not selected.

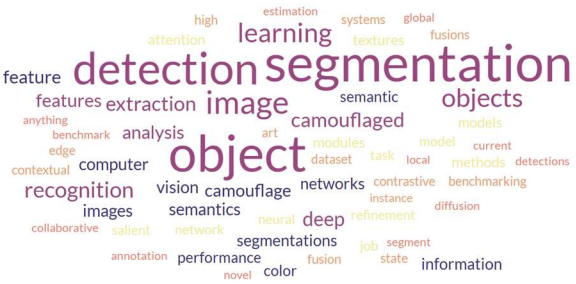

Figure 1- Keyword cloud of the included studies. This figure shows the most common terms identified from the collection of included papers, reflecting themes and trends within camouflaged object segmentation literature.

- **Type of publication:** Peer-reviewed journal publications and conference proceedings with adequate methodology.
- **Language:** Published in English.
- **Full-text retrievability:** The full text can be retrieved (through institution access, subscription, or upload by the authors) to facilitate the reliable extraction of data (architecture, datasets, metrics, and results).
- **Time Window:** Published between the years 2020 and 2026 inclusive.

iii) Exclusion Criteria:

Studies were excluded if any of the following applied:

- **Not segmentation-centered:** Tasks such as camouflaged object detection or classification that do not involve pixel-level masks.
- **Outside camouflage scope:** Research papers involving adverse image conditions (low light, underwater, hazy), unless they explicitly formulate a camouflage or camouflage-like segmentation task in the paper, are excluded from the scope of camouflage research papers.
- **Non-primary literature includes:** Reviews/Surveys, Editorials, Thesis/Dissertations, Patents, Tutorials, and Abstracts
- **Too narrow and not transferable:** Highly application-specific works were excluded if they failed to deliver transferable methodological results or general perspectives on COS evaluation.

iv) Grouping for synthesis:

To support a coherent narrative synthesis, each included study was assigned to a primary methodological category based on its central contribution as described in the main method section and supported by ablation/analysis (e.g., boundary/structure refinement, multi-scale or staged reasoning, transformer-enabled global context with local detail recovery, diffusion/iterative refinement, auxiliary cues, or task variants). When a study combined multiple mechanisms, categorization followed the dominant driver of the reported improvement to maintain consistency across the taxonomy.

v) Information Sources

The e-search to identify related work and extract relevant publications was performed on the following online databases: Scopus, IEEE Xplore, and The Lens. The reason behind the selection of these databases is that they provide broad coverage to various refereed publications in the field of computer vision and related domains. The final search on all databases was performed on January 2026.

vi) Search strategy

A structured keyword search was conducted in Scopus, IEEE Xplore, and The Lens. We used a two-block Boolean query that combined (i) camouflage-related terms with (ii) segmentation-related terms. The same conceptual query was applied across all databases, with minor interface-specific syntax adjustments (e.g., field tags, quotation handling). The canonical Boolean query was:

(camouflage OR "camouflaged object" OR "concealed object" OR "camouflaged scene" OR "background camouflage" OR "camouflaged object detection")

AND (segmentation OR "image segmentation" OR "object segmentation" OR "semantic segmentation" OR "instance segmentation" OR "pixel-wise" OR mask OR "mask prediction")

The search was intentionally broad to capture segmentation-oriented studies that may use camouflaged object detection (COD) terminology while still reporting pixel-level masks evaluated against mask-based ground truth; accordingly, the term “camouflaged object detection” was retained in the query to reduce the risk of missing such studies. No restrictions were applied at the query level regarding specific venues, publishers, or authors to minimize selection bias. The final database-specific search strings and any interface-dependent modifications are provided in Appendix A to ensure full reproducibility.

vii) Selection Process

Records retrieved from Scopus, IEEE Xplore, and The Lens were exported, merged, and de-duplicated using DOI matching when available and normalized title matching otherwise. Title/abstract screening was conducted by a single reviewer against the predefined eligibility criteria. To minimize erroneous exclusions, records with unclear relevance at the title/abstract stage were conservatively retained for full-text assessment. To improve screening consistency and reduce selection bias, a random subset of records was re-screened after a two-week interval; disagreements were resolved by revisiting the eligibility criteria and documenting the final decision. Full-text articles were then assessed for eligibility, and all screening and eligibility decisions were recorded in a structured log for traceability. No dedicated systematic review software was used.

viii) Data Collection Process

Data extraction was conducted in a standardized manner for easier comparison among papers. For each paper that was considered for review, we extracted the following information: biblio data (title, list of authors, year of publication), objective(s) of the study, design framework (which included contributions on either the model or algorithms and the datasets employed), and evaluation results (key performance metrics and results). In cases with multiple result sections that corresponded to various experiment configurations, we selected results that were most relevant for camouflaged object segmentation and the primary evaluation experiment.

ix) Data Items

The data from each of these trials was categorized into two broad areas: outcomes, or performance-concerned data, and study characteristics (study-level variables):

x) Outcomes:

The primary task was the performance of the segmentation result of the camouflaged object on the camouflaged object segmentation benchmarks. With each of the papers, the evaluation metric in the primary comparison table was recorded, including but not limited to (if available), mIoU, F-

measure or F1-score, as well as the relevant task-specific segmentation metric (for example, MAE, S-measure, E-measure, or Weighted F-Measure).

In cases where a study included a number of results (involving a variety of datasets, for instance, or a number of protocols), we extracted all results that were germane to camouflaged object segmentation. In situations involving a variety of models, we included details about variants based upon a model that was cited as a specific methodology, for instance.

xi) Other Variables:

For the purpose of structured synthesis, we extracted the following study-level variables when possible: publication year and journal, model family/design approach (for example, transformer-based, multi-scale refinement), backbone/encoder, most important architectural components, training approach (for example, auxiliary learning, self-learning, or domain adaptation learning), and loss functions, and implementation and testing details.

Where possible, efficiency-related data was also collected, including the speed of inference, model size/parameters, memory utilization, and computational complexity (such as FLOPs), as well as any specified hardware configurations.

Missing or Ambiguous Data: There was no imputation done for missing variables. If a question was not explicitly stated in the paper, it was coded as not reported. For any ambiguous description not captured in the reading, the variable was not recorded.

xii) Study Appraisal and Reporting Bias Considerations

Given the substantial heterogeneity in COS study designs and reporting conventions, we did not apply a formal clinical-style risk-of-bias tool. Instead, we performed a structured reporting-focused appraisal targeting threats to comparability and reproducibility that are most salient in computer-vision benchmarking studies. During full-text assessment and data extraction, we recorded whether each study clearly specified the evaluated datasets and protocol, reported the primary evaluation setting and metrics, included competitive baselines, provided ablation/controlled analyses supporting key claims, and disclosed implementation resources (e.g., code or sufficiently detailed training/inference configuration) and efficiency-related information when available. These considerations were used to summarize study-level strengths and limitations (Table 2) and to contextualize confidence in cross-study comparisons during synthesis, rather than to exclude studies.

xiii) Effect measures

Because this review focuses on computer vision segmentation performance, effect measures were defined as the reported evaluation metrics of each study on the corresponding benchmark(s). Metrics such as mIoU and F-measure/F1-score (and other commonly reported segmentation metrics when applicable) were extracted and presented as reported by the original studies, without conversion to a common effect size.

xiv) Synthesis methods

We performed a narrative and table-based synthesis. Studies were categorized according to their primary contribution to methodology as conveyed through their salient ablation or main focus. Outcomes were consolidated within structured tables to enable comparison across model lines, datasets, as well as methodologies. Due to the large degree of heterogeneity across datasets, metrics, train conditions, as well as reporting conventions, a quantitative meta-analysis was deemed to be inappropriate. In this case, qualitative analysis of heterogeneity was done by identifying patterns across datasets as well as methodology categories.

3 Results

Study Selection

A structured search was conducted in January 2026 on Scopus, The Lens, and IEEE Xplore, resulting in a total of 11,951 records (Scopus: 882, The Lens: 10,747, IEEE Xplore: 322). All records were exported and combined into a single dataset. Deduplication was done via DOI matching where possible, otherwise via normalized title matching. Additionally, author overlap and publication-year consistency were also used for deduplication, resulting in 1,138 duplicates being removed. A computer-based eligibility check of all records was done via pre-defined criteria, including publication years between 2020 and 2026, publication languages restricted to English, and document types restricted to peer-reviewed journals and conferences. Additionally, non-primary literature was excluded. This resulted in 10,410 records being removed from the dataset. This left 403 records that underwent the title/abstract screening stage. Of these, 310 records were excluded due to a lack of topical relevance and/or task alignment. Specifically, papers that did not have a primary contribution that was segmentation-centered were excluded. This included papers on camouflaged object detection/localization that used bounding boxes, region proposals, or keypoints instead of masks. Additionally, papers on saliency that used coarse saliency/activation maps instead of learning and evaluating masks on pixel-level ground truth were also excluded. Furthermore, papers on adverse conditions such as low-light, underwater, or haze were excluded unless they specifically defined a camouflage or camouflage-like segmentation task.

Full-text versions were sought for the remaining 93 reports. Of these, 22 could not be obtained, with the remaining 71 full-text reports being screened for inclusion. After full-text screening, 33 studies were deemed ineligible based on predetermined reasons. The reasons for exclusion were: detection/localization only (box level, no mask-based evaluation) (15 studies), out of the scope of camouflage (adverse conditions without explicit camouflage formulation) (10 studies), and insufficient methodological description for extraction/comparison (8 studies). In the end, 38 studies fulfilled all the inclusion criteria and were included in the final qualitative analysis. The entire process, including the numbers at each step, is represented schematically in the PRISMA flowchart (Figure 2). To be transparent, all full-text studies that were excluded after the initial screening are listed in Supplementary Table S1 with the primary reason for exclusion, following the aforementioned exclusion criteria.

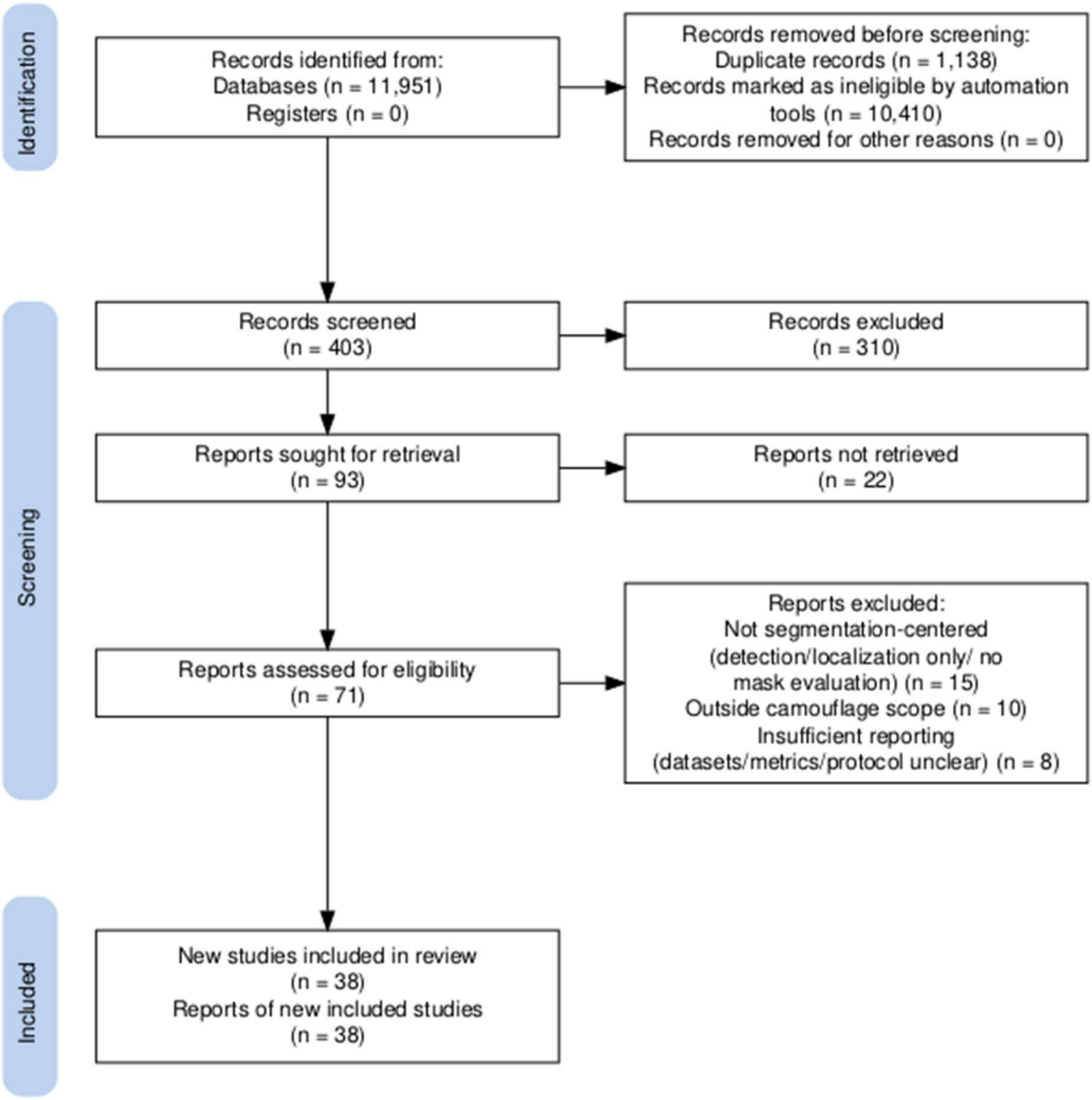

Figure 2- PRISMA Flow Diagram

Study Characteristics

The characteristics of each study were extracted in a standardized manner that facilitates direct method-focused comparison across the existing COS literature. In particular, we identified the evaluation dataset(s) and protocol adopted, the main methodological contribution identified as a core idea/category, as well as the main architectural design decisions taken, together with a description of the performance claim as represented by the main improvements identified. The characteristics are listed in Table 1, whereas a concise methodological appraisal of each study's main advantages and disadvantages with particular emphasis on transparency/repeatability and trade-offs is reported in Table 2.

Table 1- Characteristics of included studies

Study	Datasets	Core idea	Architecture	Improvements
[9]	CAMO, COD10K, NC4K	Tackles sample imbalance via cross-scale feature fusion + multi-level knowledge distillation: Feature Similarity Perception (FSP) + Gated Residual Module (GRM) to enhance sparse positive samples and suppress negative-sample interference	Encoder–decoder with shared encoder and multi-scale decoder; dual-input (original + 1.5× magnified) with weight sharing; teacher typically PVTv2-B4, students include lighter backbones (e.g., PVTv2-B2) via distillation	Reports SOTA-level gains on COD10K/NC4K; teacher (PVTv2-B4) achieves MAE = 0.017 on COD10K and claims 19.05% MAE reduction vs second-best; also reports student models reaching comparable performance to heavier backbones through KD
[10]	CAMO, COD10K, NC4K	Holistic Registration Theory-inspired global-to-local modeling: a two-stage cascaded framework (coarse semantic localization, fine detail learning). Introduces MSME (multi-scale mixture of experts) for adaptive multi-scale global priors + IFD/IFB interactive fusion for cross-scale dependency modeling during refinement	Hybrid Transformer–CNN cascade: coarse stage uses PVTv2 backbone + decoder for coarse mask; fine stage uses a lightweight CNN (Attention U-Net backbone noted) with an Interactive Fusion Decoder; trained with structure loss (weighted BCE + weighted IoU)	Claims consistent competitive/SOTA performance across the three benchmarks; e.g., reports sizable gains vs strong baselines and notable MAE reductions (e.g., vs DINet on COD10K reports ~29% MAE reduction) and best/near-best results across multiple metrics
[11]	CAMO, CHAMELEON, COD10K, NC4K	Proposes depth-surrounding-aware learning to strengthen camouflage discrimination by leveraging monocularly estimated depth and a mirrored cross-refinement strategy. Introduces dedicated modules (e.g., depth noise suppression and contour-aware enhancement) to reduce depth artifacts and improve boundary completeness.	Built on a TransNeXt backbone with early RGB-depth fusion; uses a staged refinement design with complementary branches for coarse discovery and fine exploration, coupled through interaction modules. Trained with structure-aware supervision (reported combinations of weighted BCE, weighted IoU, and Dice-style terms).	Reports consistent gains over strong baselines on major benchmarks. For example, it reports improvements on CAMO compared with a competitive baseline (ZoomNeXt), including higher structural and weighted F-measures and a notable MAE reduction; also reports advantages over prior depth-assisted COD methods.
[3]	CAMO, CHAMELEON, COD10K, NC4K	Introduces a unified "zooming in and out" mixed-scale strategy to mine subtle cues in camouflage. Key contributions include multi-head scale integration for selective multi-scale aggregation, rich granularity perception for strengthened feature discrimination, and a ground-truth-free uncertainty awareness loss to suppress	A unified image–video COD framework built around a shared triplet feature encoder operating on a 3-scale pyramid (1.0, 1.5, 0.5). Uses MHSIU modules within a scale-merging subnetwork and an RGPU-based hierarchical decoder with difference-aware conditional computation for video (activated via inter-frame differences). Trained with BCE plus the proposed UAL regularizer.	Reports state-of-the-art performance across multiple image COD benchmarks and strong gains in video COD. The paper attributes improvements primarily to mixed-scale collaboration (better localization under scale diversity), RGPU feature refinement (clearer regions and boundaries), and

		ambiguous low-confidence predictions in confusing regions.		UAL (reduced ambiguity and stronger confidence polarization).
[12]	CAMO, CHAMELEON, COD10K, NC4K	Proposes a deep unfolding formulation for COS and introduces reversible modeling across both mask and RGB domains to reduce uncertainty. Key idea is coupling segmentation (foreground separation) with reconstruction (background extraction) so that ambiguity in masks is addressed via distortion cues in RGB space.	A multi-stage unfolding network where each stage contains two reversible modules: SOFS (Segmentation-Oriented Foreground Separation) and ROBE (Reconstruction-Oriented Background Extraction). SOFS integrates Reversible State Space (RSS) for non-local refinement and includes an auxiliary edge output; ROBE uses a lightweight U-shaped reconstruction network to reconcile conflicting foreground/background estimates.	Reports leading performance on standard COD benchmarks under multiple evaluation settings (single-stage, multi-scale, multi-stage) and shows that the framework can improve existing methods when used as a refiner and can be integrated in a plug-and-play manner to strengthen baselines. Also reports robustness analyses under degraded conditions (e.g., haze simulation).
[13]	CAMO, CHAMELEON, COD10K, NC4K	Reframes COD/COS-style segmentation by isolating the environment instead of directly seeking the camouflaged object. Builds an environment prototype library (DiffPro) and retrieves background using adaptive thresholding and coarse-to-fine retrieval; the camouflaged object mask is obtained by inverting the retrieved environment.	A training-free, retrieval-based pipeline leveraging foundation models: LLaVA-1.5 for environment categories, Stable Diffusion for environment-only prototype generation, and DINO/DINOv2 as the feature extractor. Key retrieval schemes include KDE-AT (adaptive threshold), G2L (global-to-local retrieval), and SR (self-retrieval refinement). Optional integration with SAM/HQ-SAM for prompt-based segmentation evaluation.	Reports large gains over prior unsupervised baselines, including an average improvement of over 10% on COD10K for unsupervised COD. When combined with SAM, it outperforms prior prompt-based segmentation approaches and is reported to be competitive with fully supervised methods on challenging benchmarks.
[14]	CAMO, CHAMELEON, COD10K, NC4K	Proposes a prey–predator co-evolution view for COD/COS-style mask prediction. Camouflageator is an adversarial training framework that generates harder camouflaged samples to improve detector generalizability. ICEG is a detector designed to reduce incomplete masks and ambiguous boundaries using internal feature coherence and edge-guided calibration.	Camouflageator uses an auxiliary generator (reported with ResUNet backbone) trained alternately with a detector. ICEG uses an encoder–decoder with ResNet-50 as default encoder, a Camouflaged Feature Coherence (CFC) module (intra-layer and contextual aggregation plus a consistency loss), and an Edge-guided Segmentation Decoder (ESD) with edge reconstruction and edge-guided separated calibration for boundary sharpening and false prediction suppression.	Reports that ICEG outperforms prior COD detectors across the four benchmarks under multiple evaluation settings and backbones. The paper also shows Camouflageator as plug-and-play, improving several existing detectors and further boosting ICEG (ICEG+) with consistent gains reported across benchmarks.
[6]	CAMO, COD10K, NC4K	Recasts COD/COS-style mask prediction as conditional mask generation with diffusion, aiming to avoid overconfident point estimates by modeling a mask distribution and	Conditional diffusion framework with an Adaptive Transformer Conditional Network (ATCN) and a Denoising Network (DN). ATCN is PVT-based and injects the noisy mask and timestep via Zero Overlapping Embedding (ZOE) and time-token concatenation; DN is a U-shaped denoiser guided by multi-scale	Reports state-of-the-art results on the three benchmarks, including MAE = 0.019 on COD10K (ensemble variant), and claims large gains over the second-best method on COD10K (notably improved weighted F-measure

		refining masks through iterative denoising.	conditional features. Training/sampling includes SNR-based variance schedule, Structure Corruption, and Consensus Time Ensemble (CTE) (plus optional multi-sample ensemble).	and reduced MAE), attributing improvements to iterative denoising and the proposed training/sampling strategies.
[15]	CAMO, COD10K, NC4K, R2C7K	Introduces a new setup, Referring Camouflaged Object Discovery (RCOD): segment only if the referred object exists in the camouflaged image, otherwise output a blank mask. Proposes co-saliency-driven modeling to reduce false positives via two main components: CAIT (co-saliency-aware image transformations) and CSOD (co-salient object discovery with similarity-aware segmentation).	Two-stage framework. CAIT includes a Saliency Enhancement Network (SEN) (PVT encoder + layer-fusion decoding) and an Image Alignment Network using a dense correspondence module (DCM) to align the referring image to the enhanced camouflaged image. CSOD uses a Siamese/shared-weight encoder with joint channel and spatial attention, producing both a segmentation mask and a semantic similarity decision. Backbone uses PVTv2.	Reports state-of-the-art performance for Ref-COD/RCOD and substantially reduced false positives when the referred object is absent. Shows large improvements over the prior referring baseline (R2CNet) on R2C7K in Ref-COD metrics, and competitive COD performance as a by-product via SEN.
[16]	CAMO, COD10K, NC4K	Targets two persistent COD/COS failure modes, small-object miss and fine-boundary degradation, by preserving detailed spatial cues while maintaining strong global semantics. Key contributions are an Hourglass Vision Transformer for multi-scale detail retention, a Dual-Path Feature Pyramid Decoder to reduce semantic-gap dilution during fusion, and a Feature Interaction Enhancement Module to strengthen complementary low-level and high-level cues.	Transformer-based encoder–decoder. The encoder is a six-stage hourglass transformer (PVTv2-based prototype), producing paired multi-scale features. The decoder performs dual-path lateral fusion via grouped fusion modules. The FIEM uses a symmetric design with local self-attention and a GCN to enhance interaction between detailed appearance features and global semantics. Supervision uses BCE applied to multiple FIEM outputs.	Reports strong gains over recent CNN and transformer baselines across all three benchmarks, with particularly strong performance on COD10K and NC4K where small objects are frequent. For example, it reports MAE 0.020 on COD10K, and attributes improvements to hourglass detail preservation plus dual-path fusion and interaction enhancement.
[17]	CHAMELEON, CAMO, COD10K	Introduces distraction mining for COS by explicitly discovering and suppressing false-positive distractions and augmenting false-negative distractions. Uses a bio-inspired “predation” perspective with Positioning (global localization) and Focus (progressive refinement on ambiguous regions).	CNN encoder–decoder built on ResNet-50. Includes a Positioning Module with non-local channel attention and spatial attention to obtain a global coarse map, followed by stacked Focus Modules that use multi-scale Context Exploration (CE) blocks to mine distractions and refine predictions progressively. Training uses BCE+IoU for positioning and weighted BCE+weighted IoU for focus outputs.	Reports state-of-the-art performance on all three benchmarks under common COS metrics and highlights both accuracy and speed. For example, compared to SINet, PFNet reports Fwβ improvements of 7.0% (CHAMELEON), 8.9% (CAMO), and 10.9% (COD10K), and runs in real time (72 FPS).

[18]	CAMO	Bio-inspired viewpoint breaking via a mirror stream. The method assumes that horizontal flipping disrupts the “natural layout” that enables camouflage, and fuses predictions from original and flipped views to recover missed cues and improve mask completeness.	Two-stream instance segmentation framework: a main stream for the original image and a mirror stream for the horizontally flipped image, followed by mask fusion. Uses an RPN-style proposal stage with Precise RoI Pooling (PrRoI) and a joint multi-task objective (classification, box regression, mask). Supports backbones such as ResNet/ResNeXt; best-reported configuration uses ResNeXt-152.	On CAMO (camouflaged-only setting), reports strong improvements over prior methods, achieving $S\alpha = 0.785$, $E\varphi = 0.849$, $Fw\beta = 0.719$, MAE = 0.077, and shows consistent gains from the mirror stream, “augmentation in the wild,” and PrRoI pooling in ablations.
[1]	CAMO, COD10K, NC4K	Boundary-guided COD/COS-style segmentation using explicit edge semantics to improve object structure recovery. Key idea is to extract object-related edges and inject them into multi-level feature learning to strengthen boundary localization and mask completeness.	CNN encoder–decoder with Res2Net-50 backbone. Three core modules: Edge-Aware Module (EAM) to extract object-related edges from low-level and high-level features under boundary supervision; Edge-Guidance Feature Module (EFM) to fuse edge cues into features using local channel attention; Context Aggregation Module (CAM) to progressively aggregate fused features in a top-down manner for final prediction.	Reports best overall results across CAMO, COD10K, and NC4K under common COD metrics, outperforming strong baselines (e.g., JCSOD and C2FNet). The paper quantifies average gains over the second-best method as +1.80% $S\alpha$, +1.40% $E\varphi$, and +3.55% $Fw\beta$, and attributes improvements mainly to edge semantics and multi-level fusion.
[2]	CAMO, CHAMELEON, COD10K, NC4K	Enhances COD/COS-style segmentation by jointly modeling multi-scale attention refinement and explicit boundary cues. Introduces a Dual Attention Mechanism to capture scale diversity (Spotlight Attention for discriminative semantics; Modulation Attention for specific semantics), and uses a boundary branch to guide feature fusion for sharper masks.	Transformer-based encoder–decoder. Uses Swin Transformer as the multi-scale backbone. The proposed pipeline includes DAM (SAM + MAM) for feature refinement, a Boundary Extraction Module (BEM) for edge features, and a progressive Feature Aggregation Module (FAM) that fuses refined multi-scale features with edge guidance. Uses multi-branch supervision including an edge loss and a difficulty-aware style loss for attention/fusion outputs.	Reports best overall results among compared methods on the four benchmarks. The paper reports, for its final configuration, MAE 0.040 (CAMO), 0.021 (CHAMELEON), 0.021 (COD10K), and 0.030 (NC4K), with corresponding improvements in F-measure, E-measure, and S-measure, attributing gains to DAM refinement plus edge-guided aggregation.
[19]	CAMO, CHAMELEON, COD10K, NC4K	Addresses two common COD/COS segmentation bottlenecks: inaccurate localization due to limited global perception and blurred masks due to insufficient detail utilization. Proposes global localization perception plus local guidance refinement to jointly improve target positioning and	Hybrid backbone SMT-T for multi-scale feature extraction (CNN in shallow stages, transformer attention in deeper stages). A Cascade Attention Perceptron (CAP) integrates multi-scale cues using coordinate attention and spatial attention in cascaded MAM blocks. A Guided Refinement Decoder (GRD) performs top-down refinement with high-level guidance and Partial Convolution Modules (PCM) to	Outperforms 12 recent methods on the four benchmarks while emphasizing efficiency. Reports 12.74M parameters, 10.24G FLOPs, and 105 FPS real-time inference, achieving the best results in 13/16 reported metric entries and near-best in the remainder.

		boundary/detail quality while keeping the model efficient.	improve context/detail representation. Deep supervision with BCE+IoU.	
[20]	CAMO, CHAMELEON, COD10K, NC4K	Addresses ineffective use of priors by improving both prior quality and prior guidance. Proposes three complementary priors (cross-level feature prior, location prior, boundary prior) and a two-stage training strategy ("backbone supervision") to obtain higher-quality priors at low cost, then guides decoding with stage-matched priors.	Two-part framework: (1) Prior generation subnetwork using PVTv2 plus a simple convolutional decoder to produce cross-level features and priors; trained in a dedicated first stage. (2) Prior guidance via PGM modules: PGM-L uses channel self-attention for location priors (low-resolution stages) and PGM-B uses deformable convolution for boundary priors (high-resolution stages), with a top-down decoding path.	Reports SOTA/competitive results on major benchmarks. The paper highlights notable improvements such as a 12.5% MAE reduction on CAMO compared with a strong baseline (CamoFormer), and improved scores on NC4K and COD10K under common COS metrics. It also reports a generalization study showing consistent gains when applying the two-stage training strategy to several existing COS models, with minimal added cost (about ~1 MB extra parameters and 1–2 FPS drop).
[21]	CAMO, CHAMELEON, COD10K, NC4K	Re-examines COD through camouflage mechanisms (difference reduction and distraction amplification) and proposes a de-camouflaging strategy by jointly exploiting task-conflicting and task-consistent attributes between SOD and COD. Task-conflicting modeling suppresses salient distractions for better localization; task-consistent modeling aligns boundary distributions to improve mask completeness.	Two-stream encoder (SOD stream and COD stream, Res2Net-50 backbone) with task-shared decoders: a Boundary Prediction Network and an Object Segmentation Network. Key components include Region Distraction Module (RDM) and Gate Classification (GC) for distraction suppression, plus an Adversarial Learning (AL) scheme and Boundary Injection Module (BIM) for boundary enhancement and boundary-guided feature modulation. The discriminator and gated classifier are used in training and omitted at inference.	Reports best overall performance among compared methods on major benchmarks (under Fβ, E-measure, S-measure, MAE). For example, the paper reports Fβ / MAE of 0.828 / 0.063 (CAMO), 0.888 / 0.024 (CHAMELEON), 0.778 / 0.030 (COD10K), and 0.843 / 0.041 (NC4K), and shows consistent gains from RDM, GC, AL, and BIM in ablations.
[4]	CAMO, COD10K, NC4K	Breaks "visual wholeness" by matching an appropriate field of view. Uses a Visual Field Matching and Recognition Module (VFMRM) to activate candidate camouflaged regions across diverse receptive fields, then a Stepwise Refinement Module (SWRM) to progressively recover complete masks while suppressing cluttered-background interference. Also proposes an efficiency-oriented deep	Encoder–decoder with ResNet-50 or Res2Net-50 backbone. VFMRM has a local branch (grouped dilated convolutions for multi-receptive-field clues) and a global branch (multi-head self-attention for long-range semantics), fused per level. SWRM performs lightweight progressive refinement using multiplicative residual fusion between VFMRM activations and backbone "raw material" features.	Reports consistent improvements over 30 state-of-the-art models on the three benchmarks under multiple standard metrics, while maintaining real-time speed (82.6 FPS) and favorable FLOPs–accuracy trade-offs (claims faster and lighter than strong baselines such as C2FNet). Also reports effectiveness on multiple COS-related downstream tasks (e.g., polyp segmentation, lung

		supervision strategy that avoids redundant intermediate features and enables pruning during testing.		infection segmentation, road crack detection, transparent object segmentation).
[22]	CAMO, CHAMELEON, COD10K	Mimics human recognition in two steps, coarse localization followed by progressive edge/detail refinement. Uses a dedicated localization branch to produce a robust coarse prior, then an overall refinement branch that explicitly suppresses false positives and false negatives while sharpening boundaries.	Dual-branch CNN with Res2Net-50 backbone. Localization branch uses a Robust Localization Module (RLM) with ASPP and partial decoding of high-level features. Refinement branch uses an Edge Refinement Module (ERM) with context exploration and an improved SPP denoising block, plus an Attention-Guided Head (AG-Head) for spatial and channel reweighting across multiple outputs (five jointly trained predictions).	Reports strong benchmark performance, highlighting Sα = 0.827 on COD10K, with consistent gains across metrics and datasets. The paper also reports an efficiency comparison indicating 28.2 FPS with 24.1G FLOPs under its stated testing setup.
[23]	CAMO-FS	Targets the core bottleneck of camouflage under few-shot learning, weak feature separability between foreground and background. Proposes two training objectives: Instance Triplet Loss (instance-level foreground–background discrimination using RoI features) and Instance Memory Storage (class-level memory bank to stabilize and enrich class-wise camouflaged representations during fine-tuning).	Built on a two-stage Mask R-CNN/MTFA-style few-shot instance segmentation pipeline (Detectron2). Uses a standard backbone with FPN (reported experiments include ResNet-101), RoIAlign, and separate heads for classification, box regression, and mask prediction. The proposed ITL and IMS are integrated as additional losses during novel fine-tuning.	Establishes a new benchmark (CAMO-FS) and reports state-of-the-art on it. On the 5-shot setting, the proposed components substantially improve both instance segmentation and detection AP over the MTFA baseline; for example, with IMS the paper reports AP improvements reaching roughly +3.66 (segmentation AP: 5.95 to 9.61) and +4.52 (detection AP: 5.84 to 10.36) on CAMO-FS under the reported setup.
[24]	CoCOD8K	Proposes a new task, Collaborative Camouflaged Object Detection (CoCOD), where the model jointly segments co-camouflaged objects across an image group to exploit cross-image consensus cues. Also contributes CoCOD8K, a large-scale dataset designed for this group-based setting.	Bilateral-branch network (BBNet) with (i) an inter-image branch using Collaborative Feature Exploration (CFE) (feature shuffling + multi-view exploration) to mine group-level consensus semantics, (ii) an intra-image branch using Object Feature Search (OFS) to capture fine camouflage cues within a sampled image, and (iii) Local-Global Refinement (LGR) to sharpen details using both local sub-block selection and global refinement, followed by lightweight decoding.	On CoCOD8K, BBNet reports clear improvements over strong COD and CoSOD baselines. For example, compared with the second-best COD model (BSANet), it reports gains of +4.50% Sα, +5.66% Emean, +9.03% Fmean, and lower MAE; it also reports sizable improvements over the best CoSOD competitor (ICNet) on the same benchmark.
[7]	CAMO, CHAMELEON,	Proposes bilateral diffusion for COS by converting both mask prediction and depth estimation into iterative denoising problems. The depth	two-stream diffusion framework with a PVTv2 backbone for image features and two lightweight diffusion U-Nets (one for depth, one for mask). Introduces Adaptive Feature	Reports state-of-the-art results against 20 recent COS methods across four datasets, emphasizing MAE reductions. The paper

	COD10K, NC4K	branch serves as semantic guidance to reduce semantic and logical inconsistencies in segmentation, and diffusion sampling mitigates overconfidence typical of one-pass decoders.	Fusion Module (AFFM) to provide branch-specific multi-scale features and Bilateral Feature Fusion Module (BFFM) to inject depth cues into the mask branch via spatial weighting. Uses specialized training and inference strategies (variance schedule and non-linear sampling) tailored to binary-mask denoising.	reports an average MAE reduction of 4.3% and notes that BiDiCOS achieves the lowest MAE in 10 of 12 backbone-specific comparisons, while also improving depth estimation quality in joint experiments.
[25]	COD10K, NC4K, CamX	Designs COD/COS-style segmentation from a cognitive-attention perspective. Uses eye-tracking experiments to summarize attention regularities and translates them into a network that emphasizes a global–local–global search pattern, with a Bidirectional Attention Module (BAM) to simulate feedforward and feedback attention for better feature transmission and detail recovery.	ResNet-50 backbone with multi-level feature alignment. Encoder uses a residual attention module (RCAB/RIR-style residual-in-residual) and dual attention blocks, specifically position attention (PAM) and channel attention (CAM). Decoder adopts a DenseASPP-style multi-scale aggregation, and BAM forms a closed-loop by computing reverse attention (from the prediction complement) and feeding it back via short skip connections.	Reports consistent gains over selected baselines (SINet, SINet-V2, LSR, JSCOD). Reported results include COD10K: Sα 0.820, Fmean 0.827, Emean 0.879, MAE 0.032; NC4K: Sα 0.827, Fmean 0.793, Emean 0.865, MAE 0.055; CamX: Sα 0.741, Fmean 0.639, Emean 0.783, MAE 0.054. Also introduces CamX to evaluate generalization across natural and artificial camouflage.
[26]	COD10K, NC4K	Addresses Camouflage Instance Segmentation by strengthening contextual semantics in high-uncertainty regions to reduce false positives. Key ideas are mixed-scale context modeling via Feature Aggregation Module (FAM), complementary foreground and background masked attention in Global Refinement Cross-Attention (GCA), and a clustering-inspired query update mechanism, Global-Shift Self-Attention (GSA), to mitigate intra-class ambiguity and distribution shift.	Mask2Former-style CIS framework with three main parts: backbone (e.g., ResNet-50), pixel decoder (FPN-based) for high-resolution per-pixel embeddings, and a transformer decoder enhanced with FAM, GCA, and GSA. Uses multi-scale inputs (global/local views) and trains with focal and dice losses for masks plus cross-entropy for class scores.	Reports state-of-the-art CIS performance on both COD10K-Test and NC4K. The paper reports improvements over a strong CIS baseline (DCNet), including approximately +3.0% in AP50 and AP75 on COD10K-Test, and +3.2% AP75 on NC4K; it also reports consistent gains across different backbones (ResNet and Swin variants).
[27]	CAMO, CHAMELEON, COD10K, NC4K	Reduces missed detections and false alarms by introducing contrastive learning that explicitly separates camouflaged-object representations from both background and salient/non-camouflaged objects. Uses joint salient-object cues to construct positive/negative pairs	Two-stage cascaded design: MNet (feature extraction encoder + feature fusion decoder) with an Edge Guidance Module (EGM) for boundary-aware learning, followed by CNet that forms foreground/background class activation maps and performs contrastive learning with a Global Relationship Capture Module (GRCM) (batch pooling plus 1D conv) to strengthen global feature relationships.	Reports strong benchmark performance and highlights balanced precision–recall improvements. Claims average gains over suboptimal competitors of +1.93% Fm and +2.7% Fw m across datasets, and reports competitive results on all four benchmarks (e.g., NC4K

		and improve discrimination under varying camouflage strength.	Backbone uses a Swin Transformer variant (Swin-S reported as a practical default).	MAE = 0.035 with top-ranked scores in multiple metrics). Also reports a "win-win" transfer, performing strongly on salient object segmentation benchmarks.
[28]	CHAMELEON, COD10K, NC4K	Mitigates background distractors by mimicking human multi-view inspection. Generates multiple "views" via simple transformations (angle flips and close-distance resizes), then compares and fuses multi-view features to enhance both boundary cues and semantics. Key designs are CAMV (two-stage co-attention across views) and CFU (iterative channel-wise contextual cue mining).	Shared-weight ResNet-50 + FPN encoder applied to multiple transformed views, followed by a view-combining layer (channel concatenation per pyramid level). Multi-view fusion is done by CAMV modules at each FPN level, then decoded with a hierarchical channel-fusion decoder using CFU blocks. Uses BCEL + UAL (uncertainty-aware loss) for training stability under camouflage ambiguity.	Reports best overall performance on all three test sets under common COD metrics. For example, it reports COD10K: Sm 0.846, Fwβ 0.745, MAE 0.028 and NC4K: Sm 0.856, Fwβ 0.791, MAE 0.042. It also reports improvements over ZoomNet on COD10K and NC4K across Sm/Fwβ/Fβ/Em, attributing gains to multi-view fusion (CAMV) and channel-wise cue mining (CFU).
[8]	CAMO, COD10K, NC4K	Proposes a two-stage segmentation–diffusion framework inspired by coarse-to-fine human perception. Stage 1 uses diffusion to reduce uncertainty in mask estimation and generate a robust prior; Stage 2 performs semantic refinement with contextual aggregation to sharpen local structure and boundaries.	PVTv2-b4 backbone with a diffusion branch and a segmentation refinement branch. Key modules: FEM (Feature Emphasis Module) for multi-scale feature expansion and redundancy suppression prior to diffusion refinement; Diffusion U-Net for conditional denoising; RM (Refine Module) for multi-scale contextual aggregation and differential feature interaction between diffusion-refined representations and local segmentation cues. Inference averages multiple diffusion-step predictions (default 10 steps).	Reports consistent gains over 17 recent methods across all three benchmarks. For example, on COD10K it reports Sα 0.883, Fwβ 0.817, Em 0.944, MAE 0.019, and states improvements over the second-best (SDRNet) of +1.2% Sα and +3.2% Fwβ on COD10K; also reports strong results on CAMO and NC4K.
[29]	CAMO, CHAMELEON, COD10K, NC4K	Leverages global–foreground visual consistency as a supervisory signal to handle extreme object–background similarity. The method first extracts a foreground-biased representation, then explicitly evaluates consistency between foreground and global context to derive a consistency attention map, which guides multi-scale refinement for sharper boundaries and fewer distractions.	Coarse-to-fine pipeline with three modules: Primary Detection Module (PDM) to fuse multi-scale backbone features and produce a coarse mask, plus a learnable-threshold filter to generate a foreground feature; Consistency Evaluation Module (CEM) that processes foreground and global features with shared-parameter dual discriminators (GAN-inspired) to generate consistency attention; Detail Refinement Module (DRM) that injects the attention into hierarchical feature fusion (top-down) using residual attention blocks and	Reports consistent SOTA performance across metrics. For example, with EfficientNet-B4, the paper reports CAMO-test Sα 0.845, Fwβ 0.793, MAE 0.052, CHAMELEON MAE 0.021, and NC4K Sα 0.868, Fwβ 0.811, MAE 0.038; with PVT, it reports further improvements including NC4K MAE 0.029. Ablations show measurable gains from the filter, shared discriminators, and deconvolution upsampling (largest reported Fwβ gain up to

			deconvolution upsampling. Reports backbone variants including EfficientNet-B4 and PVT.	+0.024 on CAMO when switching to deconvolution).
[30]	CAMO, COD10K, NC4K	Proposes an implicit feature selection strategy to reduce background interference and a multi-scale region attention strategy to enhance target regions. The key idea is to jointly exploit local CNN cues and global transformer cues, then retain only the most contributive feature channels and reinforce multi-scale target consistency.	Hybrid dual-backbone design using Res2Net-50 for local features and PVTv2-b1 for global features, followed by Implicit Feature Selection (IFS) (channel-weighted selection using ECA and top-half retention) and stacked Multi-scale Region Attention (MRA) modules; multi-level fusion is implemented with an FPN-style decoder and trained with weighted BCE, weighted IoU, plus uncertainty-aware loss (UAL).	Reports best overall performance among 20 compared methods on all three datasets under Sα, Fwβ, MAE, and Em. It reports improvements over the second-ranked method such as: on CAMO, about +1.5% Sα and +1.6% Fwβ with lower MAE; on COD10K, about +0.7% Sα and +0.5% Fwβ; on NC4K, about +1.8% Sα, +2.2% Fwβ, and a notable MAE reduction (reported as 9.3% relative decrease).
[31]	CAMO, CHAMELEON, COD10K	Mimics a two-stage human visual observation process for camouflage. Stage 1 performs coarse search; Stage 2 refines using dual guidance that combines holistic attention and reverse attention. A final refinement stage leverages peer-layer feature cues to improve completeness and suppress distractors.	CNN encoder–decoder with Res2Net backbone. Main components: DFE (Dual-branch Feature Extraction) with a two-stage design and RF blocks; HAREW module for holistic-attention plus reverse-attention guidance and weighted fusion; GRCF (Gradually Refined Cross Fusion) using a U-Net-like refinement with Self-Refine Attention (SRA) and Cross-Refinement (CR) units. Trained with weighted BCE plus weighted IoU and deep supervision across stages.	Reports SOTA on all three datasets using standard metrics. Example headline results: COD10K Fmean = 0.720 and MAE = 0.037; the paper states COD10K Fmean is +0.086 (13.56%) higher than the second-ranked model and MAE is 0.014 (27.45%) lower.
[32]	MAS3K	Targets underwater segmentation under two coupled difficulties: water degradation diversity and camouflage-like appearance of marine animals. Introduces Random Style Adaption (RSA) and a siamese consistency objective to reduce sensitivity to low-frequency style shifts, while enforcing consistent predictions across original and RSA-augmented views.	Siamese training framework with a shared MainNet built on an asymmetric encoder–decoder using Res2Net-50 features. Uses Receptive Field Blocks (RFB) for multi-scale feature enrichment and Attention Fusion Blocks (AFB) (SimAM-based) for multi-level contextual fusion. Training uses BCE + SSIM + IoU loss for each output plus a consistency loss between siamese outputs; RSA is removed at test time.	On MAS3K-Test, reports best performance among compared SOD/COD baselines (e.g., SINet, SINet-V2, C2F-Net, SCRN, BASNet), achieving mIoU 0.739, Sα 0.856, Fwβ 0.804, mEφ 0.913, MAE 0.032. The paper reports measurable gains from each component (RSA, RFB, AFB) in ablations.
[33]	CAMO, CHAMELEON, COD10K, NC4K	Improves COD/COS-style segmentation by strengthening feature discriminability and context modeling, then performing redundancy-aware cross-level aggregation. Key modules are	CNN encoder–decoder with Res2Net-50 backbone. Uses triplet mixed-scale inputs (1.0×, 1.5×, 0.5×) with per-level DIAM blocks; CEM fuses adjacent levels using multi-branch dilated convolutions; CFAM aggregates low- and high-level features using element-wise	Reports best results among compared methods on all four benchmarks. For example, it reports COD10K MAE = 0.028 and NC4K MAE = 0.041, with stated improvements over the

		DIAM (mixed-scale discriminative information attention), CEM (context enrichment via multi-rate dilated convolutions), and CFAM (cross-level feature aggregation with multiplicative fusion to suppress background distractors).	multiplication plus concatenation in a top-down decoding process. Trained with weighted BCE + weighted IoU and deep supervision.	second-best method on COD10K of +1.7% Sα, +3.4% Eξ, +2.3% Fβ, and 0.4% MAE improvement (absolute).
[5]	CAMO, CHAMELEON, COD10K, NC4K	Bridges the gap between global localization (strong in Transformers) and fine textures/boundaries (strong in CNNs) by explicitly designing interaction between the two representations. Introduces Texture Feature Fusion Module (TFFM) to softly combine CNN texture cues with shallow transformer structure cues, and Noise Removal Module (NRM) to suppress low-level background noise using transformer-derived localization priors.	Dual-encoder framework with Res2Net-50 (first three stages) and CoaT-Lite transformer encoder. TFFM fuses CNN features with shallow transformer features using learned soft-selection weights (channel and spatial attention-guided fusion). Transformer high-level stages provide global positioning cues. NRM enhances foreground–background contrast using sigmoid-activated positioning features and a texture enhancement module, followed by attention-based refinement. Training uses weighted BCE + weighted IoU.	Reports best results versus 7 strong COS baselines across all four datasets. Example headline scores: CHAMELEON MAE 0.020, CAMO MAE 0.047, COD10K MAE 0.024, NC4K MAE 0.033. The paper highlights improvements on COD10K over UR-SINet-v2 of +4.7% Sα, +1.6% Eadφ, +7.1% Fwβ, and 0.9% lower MAE (absolute).
[34]	CHAMELEON, CAMO, COD10K, CPD1K	Argues that contrast cues alone are insufficient for camouflage, and introduces part–object relational knowledge as a complementary signal to improve object completeness and reduce missing parts and boundary failures. Proposes a two-stage search–identification pipeline where contrast and part–whole cues are explicitly integrated during decoding, and the search stage guides the identification stage.	Two-stage encoder–decoder. Each stage contains a Contrast Information Exploration (CIE) encoder and a Part-Object Relationship Exploration (PORE) encoder implemented via a CapsNet (matrix capsule routing) for part–whole modeling. Decoding uses POGU (Part-Object relationship Guidance Upsampling) to integrate contrast features and capsule-based relational cues via feature combination, channel-wise promotion, and POR guidance. Stages are connected by SIG (Search-to-Identification Guidance), which injects the search prediction and decoded search features into the identification-stage encoder to improve feature encoding and suppress noise. Backbone for CIE uses VGG16 feature stages (Conv blocks).	Reports strong gains on multiple benchmarks, with the most pronounced improvement on CPD1K (camouflaged people), where it reports increasing Fβ over the prior best method by roughly ~17 points. It also reports competitive overall averages across datasets and an efficiency note of about 0.1 s per 352×352 image under the stated setup.
[35]	CAMO, CHAMELEON,	Argues that conventional context modeling is insufficient because it ignores spatial positional relationships among objects in the scene. Proposes a spatial perception	CNN encoder–decoder with Res2Net backbone. Three modules: PAM (Perceptual Activation Module) builds positional relationships using GCN with self-attention and reversed self-attention adjacency; FIM (Feature	Reports competitive results against 13 recent methods on four benchmarks. Reported headline metrics include: CHAMELEON (Eφ 0.950, Sα 0.901, Fwβ 0.859,

	COD10K, NC4K	strategy to model object–environment positional relations for better localization in cluttered backgrounds, and fuses features by jointly mining semantic correlation and detail discontinuity to reduce over- and under-segmentation.	Inference Module) uses correlation information to suppress background interference and re-encode multi-scale features; IRM (Interaction Recovery Module) fuses adjacent scales while explicitly mining local and global discontinuity (5×5 conv and global pooling branches) to sharpen structure.	MAE 0.026), CAMO (Eφ 0.883, Sα 0.822, Fwβ 0.772, MAE 0.070), COD10K (Eφ 0.905, Sα 0.835, Fwβ 0.727, MAE 0.031), NC4K (Eφ 0.906, Sα 0.849, Fwβ 0.788, MAE 0.045).
[36]	CAMO, CHAMELEON, COD10K, NC4K	Recasts COD/COS-style mask prediction as a denoising diffusion process from noisy masks to object masks, aiming to improve fine-grained texture and boundary segmentation by iterative refinement rather than one-shot deterministic decoding.	Diffusion framework with a UNet-based denoising network conditioned on image features. Uses a PVTv2 backbone and a Feature Fusion (FF) module to build multi-scale conditional semantics. Introduces an Injection Attention Module (IAM) (cross-attention) to inject conditional image semantics into diffusion features for stronger denoising guidance. Uses a diffusion schedule with T = 1000 and trains with a hybrid objective combining simplified diffusion loss, VLB term, and a static-mask auxiliary loss.	Reports favorable performance versus 11 recent COD methods on four benchmarks. Example results from the main table include COD10K MAE = 0.036, NC4K MAE = 0.051, CAMO MAE = 0.082, CHAMELEON MAE = 0.030, with top-ranked scores in multiple metrics. The paper also reports notable average relative gains, including 19.1% MAE reduction on COD10K and 14.8% MAE reduction on NC4K compared with prior methods under the stated setup.
[37]	CAMO, CHAMELEON, COD10K, NC4K	Addresses the two dominant COD/COS failure modes, low contrast and high interference, with a biologically inspired three-stage pipeline: Search builds complementary representations (RGB plus frequency cues), Scan enforces region-level spatial coherence while accounting for ambiguity, and Recalibration performs progressive reasoning to recover thin structures and stabilize boundaries.	Transformer-centric framework using PVTv2 backbone with dual-domain encoding (RGB and DCT-derived frequency). The Region-aware Scan Module (RSM) introduces a CRF-inspired energy formulation with window-restricted block-level pairwise interactions for efficient spatial consistency. An Uncertainty Estimation Module (UEM) injects pixel-wise uncertainty (sampling-based variance) to gate unreliable regions. A Cognitive Progression Module (CPM) performs iterative feedback refinement (multiple iterations with intermediate supervision) to progressively sharpen predictions.	Reports top-ranked results on all four benchmarks. The paper reports CAMO: Sα 0.868, Fwβ 0.822, MAE 0.049; CHAMELEON: Fwβ 0.873, MAE 0.021; COD10K: Sα 0.861, Fwβ 0.770, MAE 0.025; NC4K: Sα 0.885, Fwβ 0.837, MAE 0.033. It also reports sizeable gains over recent baselines (e.g., improvements on CAMO weighted F-measure over DSNet, and improved robustness under LC/HI), and provides efficiency indicators (Params, FPS, runtime) to contextualize accuracy–cost trade-offs.
[38]	CAMO, COD10K, NC4K	Two-stage COD/COS-style segmentation inspired by hierarchical human perception: stage 1 focuses on edge and position	PVTv2-b4 backbone with three key modules: Edge Exploration Module (EEM) plus Retrieve Attention for refined boundary cues; Object Position Recognition Module (OPRM) to detect	Reports outperforming 16 recent methods on COD10K, NC4K, CAMO. Best reported backbone setting (PVTv2-b4) achieves

	discovery, stage 2 strengthens contextual aggregation for robust mask prediction.	horizontal/vertical position signals and enhance features via multi-dilation expansion and neighbor fusion; Context Aggregation Module (CAM) for top-down multi-level fusion and final mask prediction.	COD10K MAE 0.020, NC4K MAE 0.030, CAMO MAE 0.041 with corresponding strong Sα, Fwβ, and Em. The paper states gains over the second-best on COD10K of +1.1% Sα and +2.7% Fwβ, and on NC4K of +0.5% Sα and +1.4% Fwβ under the same evaluation setting.

Table 2- Study appraisal: strengths and limitations of included studies

Study	Strengths	Limitations
[9]	Clear problem framing (sample imbalance) and targeted modules (FSP/GRM) + feature- & response-level KD; evaluation on 3 mainstream benchmarks with standard splits; code released	Primarily in-dataset benchmarking (no explicit cross-dataset generalization study); dual-input magnification and teacher backbone can increase compute (deployment cost may be non-trivial); focus is COD-style segmentation (mask prediction) rather than task variants requiring instance-level separation
[10]	Clear, well-motivated two-stage global; local design; strong ablation showing contributions of cascade structure + MSME + IFD; broad comparison against 19 SOTA methods and includes qualitative + failure case analysis	code not openly released ("available on request"); evaluation focuses on the common COD benchmarks (no explicit cross-domain/generalization study beyond these datasets); efficiency reporting is limited (no standard FPS/FLOPs table)
[11]	Clear motivation for using depth as an auxiliary cue for camouflage discrimination; explicit module design targeting depth noise and boundary/detail recovery; evaluation on multiple standard benchmarks with extensive comparisons; reports implementation availability (code release stated).	Relies on a depth-estimation pipeline, which can increase computational cost and may confound comparisons with purely RGB methods; performance may depend on the quality and domain suitability of predicted depth; evaluation is primarily benchmark-focused with limited emphasis on cross-domain robustness beyond standard datasets.
[3]	Strong methodological coherence: scale collaboration (MHSIU), feature discrimination (RGPU), and reliability-oriented optimization (UAL) are clearly motivated and empirically validated via ablations. The unified design covers both image and video COD under a single framework, reducing task-specific duplication. Extensive benchmarking and qualitative results support robustness claims.	The explicit multi-scale "triplet" processing increases computational demand compared to single-scale pipelines (the authors acknowledge additional inference cost). Performance and efficiency may depend on backbone choice and input resolution, and the method is primarily validated on established benchmarks rather than challenging cross-domain settings.
[12]	Strong theoretical motivation with an interpretable unfolding-based design; explicitly targets COS uncertainty by jointly modeling mask refinement and RGB reconstruction; extensive experiments on major COD benchmarks and multiple COS-related tasks; includes	The paper frames COS as an umbrella covering multiple tasks, which can blur strict COS scope depending on review criteria; the multi-stage unfolding design can be computationally heavier than single-pass models; at the time of the manuscript it states code will be released, so reproducibility may depend on actual availability;

	ablation studies validating key components (SOFS/ROBE/RSS, reconstruction, staging).	evaluation is primarily benchmark-driven with limited emphasis on cross-domain generalization beyond reported auxiliary analyses.
[13]	Strong conceptual novelty with a clear and testable hypothesis (environment-first separation); fully unsupervised and training-free design reduces annotation/training cost; well-structured method with targeted components (DiffPro, KDE-AT, G2L, SR) supported by ablations and comparisons; reports code availability.	Depends on multiple foundation models and prototype generation, which can add computational and engineering overhead (prototype library construction, feature extraction, retrieval at scale); outcomes may be sensitive to prototype quality and environment category coverage; evaluation is benchmark-centric, and cross-domain robustness is only partially addressed (generalization shown on a non-COD domain but not framed as systematic domain shift).
[14]	Strongly motivated two-part contribution: (i) a flexible adversarial training framework that can enhance multiple detectors, and (ii) a detector that explicitly targets two dominant failure modes (incompleteness and boundary ambiguity). Extensive benchmarking and thorough ablations support the roles of CFC, ESD, and the adversarial training procedure.	The framework relies on ground-truth masks for training the generator and thus does not reduce annotation needs; the adversarial sample generation introduces additional training complexity and may create distribution shift that is only partially characterized. The method is benchmark-driven and does not provide a dedicated cross-domain evaluation beyond standard datasets. Code availability is stated as forthcoming.
[6]	Strong conceptual shift from deterministic segmentation to probabilistic mask generation; method design is well-scaffolded (ATCN/DN plus clearly isolated training and sampling strategies) with extensive ablations; broad comparison against many recent baselines on standard benchmarks; public repository is provided.	Inference can be computationally heavier than single-pass models due to iterative denoising and optional ensembling; results can be sensitive to sampling settings (steps, seeds, ensemble size), which complicates deployment; evaluation is primarily benchmark-focused on common COD datasets, with limited emphasis on systematic cross-domain robustness.
[15]	Clearly defined task formulation (RCOD) that directly targets a real failure mode of generic COD, false positives under specified-object search; well-structured pipeline (CAIT + CSOD) with explicit similarity prediction; introduces Ref-1K to broaden referring coverage; includes cross-domain style analysis within the referring setting; code is released.	Not a standard single-image COS/COD setting, it requires a suitable referring image and additional transformation/alignment steps; performance depends on the quality of saliency enhancement and correspondence alignment; evaluation and metrics require careful handling of negative (blank-mask) cases, which makes direct comparison to conventional COS/COD studies less straightforward.
[16]	Clear, well-motivated architectural contributions aligned with stated failure modes; extensive comparison against a large set of strong baselines; solid ablation studies isolating the hourglass backbone, decoder, and FIEM effects; reports implementation details and computational indicators (Params and GFLOPs) and provides a public code repository.	Computationally heavy for deployment in its strongest configuration (the full model reports high GFLOPs, and training uses high-end GPU hardware), which may limit real-time or edge use; evaluation is mainly benchmark-driven on standard COD datasets with limited emphasis on systematic cross-domain generalization or robustness beyond the benchmarks.
[17]	Strongly aligned design with COS failure modes by modeling both false positives and false negatives; clear modular decomposition (Positioning + Focus) with multi-scale context reasoning and progressive refinement; extensive comparisons against a diverse set of baselines and	Evaluation is limited to the standard three COS benchmarks without a dedicated cross-domain/generalization protocol; uses a conventional CNN backbone (ResNet-50), so comparisons to later transformer-era models may not reflect current best practice; the multi-module refinement pipeline adds architectural complexity relative to simpler single-pass designs.

	solid ablations; emphasizes practical efficiency with real-time throughput.	
[18]	Clear and interpretable idea (dual-view inference with mirror fusion) with ablations validating mirror stream, data augmentation, and pooling choice; explicitly addresses data scarcity via "augmentation in the wild"; reports consistent metric gains and qualitative improvements on challenging CAMO cases.	Evaluation is primarily CAMO-only, which limits conclusions about generalization to later large-scale benchmarks (e.g., COD10K/NC4K); relies on an instance-segmentation-style pipeline (proposal, RoI operations, fusion) that can add engineering and compute overhead compared with single-pass fully convolutional models; code availability is stated as planned via project pages, so reproducibility depends on actual release status.
[1]	Clear alignment between method design and the stated failure mode (boundary incompleteness). Strong ablations isolate the impact of EAM, EFM, and CAM; competitive efficiency reporting is provided (Params, FLOPs, FPS) alongside multi-task COD baselines; code is publicly available.	Uses a CNN backbone and module-based design that may be less competitive against later transformer-centric paradigms under the same compute budget; evaluation focuses on standard COD benchmarks with no dedicated cross-domain generalization protocol; model size and compute are non-trivial compared with lightweight designs.
[2]	Clear, modular contributions aligned with common COD/COS failure modes (scale diversity and weak boundaries). Includes ablations validating SAM/MAM, boundary extraction, and edge-guided aggregation. Benchmarked on four standard datasets with comparisons to multiple recent baselines.	Evidence is limited to benchmark datasets with no dedicated cross-domain generalization protocol. The method relies on a transformer backbone and multiple refinement branches, which can increase training and inference cost; the paper does not provide a consolidated efficiency table (FPS/FLOPs) for direct deployment comparison. Code availability is not clearly stated in the manuscript.
[19]	Strong efficiency focus with explicit reporting of Params, FLOPs, and FPS; method components are tightly aligned with stated failure modes (CAP for multi-scale global localization, GRD/PCM for boundary and detail refinement); thorough benchmarking against 12 competitors with ablations validating CAP and GRD contributions; public code link is provided.	Benchmark-centric evaluation on standard datasets with no dedicated cross-domain generalization protocol; improvements may partially depend on the specific hybrid backbone choice (SMT-T) and resolution settings; failure cases remain for heavy occlusion and multiple camouflaged objects, as acknowledged by the authors.
[20]	Strong methodological clarity: separates "prior quality" from "prior usage" and designs priors matched to decoding stages; the two-stage training ("backbone supervision") is well-motivated and supported by ablations; includes a useful transferability/generalization experiment by applying the training strategy to other COS models; provides a public code repository.	Evaluation remains largely benchmark-centric on standard COS datasets; improvements are partly tied to transformer backbones (PVTv2/ConvNeXt variants), so compute budget can still be non-trivial; while transferability is demonstrated across models, cross-domain robustness beyond the benchmark suite is not framed as a dedicated domain-shift protocol.
[21]	Strong conceptual framing grounded in camouflage mechanisms, with a clear mapping from failure modes to modules (distraction suppression and boundary completion). Thorough ablations isolate contributions of RDM, GC, AL, and BIM, and the paper also reports complexity indicators (Params, FLOPs, FPS) under a fixed input size for fair comparison. Inference-time design is streamlined by	Requires additional SOD data and labels during training, which increases supervision requirements and training complexity relative to COD-only methods. The full framework is architecturally complex (two-stream encoding plus multiple auxiliary modules), and the authors note the performance gain comes with increased complexity that can limit deployment practicality. Evaluation is

	removing training-only components (discriminator and gated classifier).	benchmark-centric, with limited emphasis on systematic domain-shift generalization beyond the standard dataset suite.
[4]	Clear biological/vision-inspired motivation translated into concrete modules (VFMRM, SWRM) with strong ablations supporting each component and the supervision design. Emphasizes practical deployment with explicit efficiency evidence, including FPS and FLOPs comparisons, and provides a public code repository. Includes downstream task evaluations that strengthen applicability beyond core benchmarks.	Evaluated primarily on the standard benchmark suite, with limited emphasis on a dedicated cross-domain generalization protocol outside those datasets. The "field-of-view matching" design is implemented via multiple receptive-field branches and attention, which can complicate architectural simplicity compared to single-pass lightweight baselines; performance may depend on backbone choice and input resizing settings.
[22]	Clear, defensible decomposition of the task into localization and refinement, with modules that directly target typical COD/COS failure patterns (imprecise localization, blurred edges, FP/FN noise). Provides ablations for key design choices (RLM feature sets, ASPP, ERM depth, SPP integration, AG-Head) and includes explicit efficiency reporting (FPS and FLOPs).	Evaluation is limited to three common datasets and does not provide a dedicated cross-domain generalization protocol. The overall design uses multiple prediction heads and refinement stages, which increases training complexity. Data availability is stated as "available on demand," and the paper does not clearly state public code release, which can affect reproducibility.
[23]	Introduces CAMO-FS, a rare benchmark explicitly designed for few-shot camouflaged animal detection and instance segmentation. The method contribution is focused and defensible (two losses directly targeting foreground–background separability at instance and class levels), supported by ablations across shot counts and backbone/base-model settings. Code is released in a public repository.	Not a standard single-image COS setting; it is few-shot detection + instance segmentation, requiring instance-level supervision and a two-stage fine-tuning protocol. Reported AP values are still relatively low in absolute terms (reflecting task difficulty and dataset properties), and conclusions are primarily tied to CAMO-FS rather than broad cross-dataset generalization. The approach inherits the complexity and compute overhead of Mask R-CNN-style pipelines.
[24]	Defines a well-motivated new setting (group-based camouflage segmentation) and backs it with a large-scale annotated dataset (CoCOD8K, 8,528 images; 5 super-classes and 70 sub-classes) and a comprehensive benchmark against 18 existing models. The baseline is modular and interpretable (CFE, OFS, LGR) with strong ablations showing each module's contribution.	Not the standard single-image COS/COD setup, it assumes access to relevant image groups, so applicability depends on how groups are formed at inference time. Performance can be sensitive to group composition and the "same-property" assumption. The dataset is constructed by reorganizing existing COD datasets, which may inherit biases and does not directly evaluate cross-domain robustness beyond the new CoCOD benchmark.
[7]	Strong conceptual contribution: reframes COS as iterative denoising rather than direct decoding, and couples segmentation with depth estimation in a principled bilateral design. Clear ablations validate the depth branch, BFFM, AFFM, and the training/inference strategies. Evaluated against a large comparison set (20 methods) on four benchmarks, and provides a public repository link for code.	Joint diffusion inference introduces additional computational overhead compared with single-pass models (iterative sampling). Depth guidance is not uniformly reliable; the paper explicitly notes cases where depth estimation is ineffective and can degrade results. Evaluation remains benchmark-centric with a standard split, without a dedicated cross-domain protocol beyond the four datasets.
[25]	Adds interpretability by grounding the method in eye-tracking-based cognitive rules and mapping them to explicit modules (PAM/CAM, RIR, BAM). Introduces a new dataset	Benchmarking is limited to a small set of baselines compared with the broader SOTA landscape in recent COD/COS literature. The paper itself acknowledges that generalization can be further

	(CamX) containing both natural and artificial camouflage for broader evaluation. Provides ablations showing the contribution of BAM and other components.	improved due to limited sample diversity and that the cognitive experiment lacks strict variable control, resulting in more qualitative rules without strong quantitative support. Code availability is not clearly specified.
[26]	Clear task-specific diagnosis and solution design for CIS, especially the complementary foreground/background masked attention that directly targets false positives under camouflage. Strong quantitative evidence with comparisons across two benchmarks and thorough ablations for FAM, GCA, GSA, and input-scale choices. Public code repository is provided.	Focused on instance segmentation rather than standard binary COS, so direct comparability depends on whether CIS is within scope. Uses multi-scale inputs and transformer decoding, which increases computation; the paper reports a higher GFLOPs cost as a trade-off. Evaluation is limited to COD10K and NC4K for CIS, with no broader cross-domain protocol beyond these datasets.
[27]	Clear failure-mode targeting (misses and false alarms) with a method that directly optimizes feature separability via contrastive learning. The design is modular and supported by comprehensive ablations (components and backbone size) and comparisons against a broad set of recent COS/COD methods. Adds boundary sensitivity through explicit edge guidance, and provides a public repository for code and evaluation resources.	Requires paired salient-object samples and additional contrastive machinery, which increases training complexity relative to standard COS pipelines. Strong performance is reported under controlled backbone/input-size fairness rules (excluding some high-capacity/high-resolution competitors), which can affect comparability to the absolute best reported numbers in the broader literature. The “joint SOS” angle may be considered outside strict COS-only designs depending on review scope.
[28]	Strongly motivated design that directly targets single-view COD weaknesses (distractors and fuzzy boundaries) using a simple but effective multi-view strategy. Provides detailed ablations on view combinations, CAMV (one-stage vs two-stage attention), and CFU contributions. Uses the same training data as competitors and reports comprehensive benchmark comparisons (including PR and Fβ curves).	Multi-view inference increases input processing and can raise test-time cost compared with single-pass models; performance depends on the chosen set of view transformations and their hyperparameters. Evaluation is primarily within standard benchmarks (limited evidence for systematic domain-shift robustness). Code availability is stated as planned (“will be available”), so reproducibility depends on actual release status.
[8]	Clear and defensible hybrid design combining diffusion robustness with segmentation precision; strong ablations quantify the contributions of FEM, RM, and the two-stage formulation; includes controlled analyses for diffusion-step selection and reports resource indicators (e.g., Params, FLOPs, FPS) enabling a transparency check; provides additional evaluations beyond COD (polyp segmentation and industrial defect examples) to support adaptability claims.	Computationally heavy for deployment, with reported high model size and FLOPs and low FPS under its stated setup; evaluation is benchmark-centric on the common three datasets and does not include a dedicated cross-domain protocol (auxiliary tasks are shown but not framed as a systematic domain-shift study); data availability is “on request,” and the paper does not clearly state an open-source code release, which can affect reproducibility.
[29]	Strongly aligned design: the paper operationalizes “camouflage as consistency” and turns it into a measurable attention signal via global–foreground consistency estimation. Provides extensive ablations on PDM/CEM/DRM, the filter, shared discriminators, and upsampling choices, making the causal contribution defensible. Uses the standard multi-dataset benchmark	The discriminator-style consistency module and deconvolution-based decoding can be computationally heavy at larger channel sizes (the paper reports very large FLOPs/parameter counts at its strongest setting). Results are still primarily benchmark-centric; while additional labeling (localization/ranking) and interdisciplinary experiments are explored, the work does not formalize a dedicated domain-shift generalization protocol beyond the standard dataset suite.

	protocol and reports both CNN and transformer backbone instantiations for broader comparability.	
[30]	Clear and defensible design rationale that directly targets camouflage failure modes (background interference and missing fine details) using two focused modules (IFS and MRA). Strong empirical evidence with comparisons against a large set of baselines and ablations isolating IFS and MRA contributions. The hybrid CNN–Transformer setup is explicitly described and reproducible at the method level.	Evidence is limited to in-benchmark evaluations on three standard datasets, with no dedicated cross-domain robustness protocol. The method depends on a dual-backbone feature extractor, which increases implementation and compute complexity, and the channel-pruning choice in IFS may be sensitive to hyperparameters and backbone scaling. Public code availability is not clearly stated in the manuscript, which may affect reproducibility in practice.
[31]	Clear, interpretable design aligned with camouflage failure modes via staged refinement and explicit forward/reverse guidance. Provides ablations showing monotonic gains as modules are added (backbone, Stage-1, Stage-2 with guidance, then GRCF). Reports both accuracy and an efficiency comparison versus SINet (parameters and FPS).	Benchmark coverage is limited to three datasets (CAMO, CHAMELEON, COD10K) and does not include a dedicated cross-domain generalization protocol. The model is slower and heavier than the strongest baseline used for efficiency comparison (SINet) under the reported setup. Code availability is stated as forthcoming, so reproducibility depends on the actual release.
[32]	Strong task-aware design for underwater settings: explicitly addresses degradation-style diversity with RSA and enforces invariance via siamese consistency. Clear modularity (RFB, AFB, RSA) with ablations demonstrating incremental benefits. Reports improvements over multiple strong SOD/COD baselines on MAS3K with consistent gains across five metrics.	This is underwater object segmentation, not standard natural-image COS, so inclusion depends on whether your SLR scope allows "camouflage-like" segmentation under underwater degradation. Evaluation is centered on a single domain dataset (MAS3K), limiting conclusions about general COS benchmarks. Code is stated as "will be available," so reproducibility depends on the actual release status.
[33]	Well-structured module decomposition with a clear mapping from failure modes to design choices (discriminability, context enrichment, and cross-level aggregation). Provides quantitative comparisons across four benchmarks and includes ablations isolating CFAM, CEM, and DIAM effects.	Uses mixed-scale processing and multiple per-level attention modules, which increases parameters (the authors explicitly note the model requires more parameters and aims to reduce them in future work). Efficiency indicators are not consolidated into a standard FPS/FLOPs table for easy deployment comparison, and evaluation remains benchmark-centric without a dedicated cross-domain protocol.
[5]	Strong and publication-defensible motivation with a clean mapping from COS failure modes to modules: TFFM for complementary texture–structure fusion and NRM for noise suppression under weak boundaries. Ablations directly verify the benefit of CNN+Transformer cooperation and isolate the effect of TFFM and NRM. Evaluation follows a widely used training protocol (CAMO+COD10K) and reports comprehensive results on four benchmarks.	Evidence remains benchmark-centric with no dedicated cross-domain robustness protocol. Uses a dual-backbone design, which increases engineering complexity and may raise compute/memory relative to single-backbone models; the paper does not consolidate deployment-oriented efficiency indicators (Params/FLOPs/FPS) alongside the main comparisons, making practical trade-offs harder to audit.
[34]	Strongly motivated by a concrete diagnosis (contrast-only misses parts, part-relational alone yields blurry boundaries/holes) and validates this with targeted design. The model is modular and well-supported by ablations for POGU, SIG, and the part–object decoding strategy, and	Uses a VGG-based contrast encoder and a CapsNet routing branch, which can limit scalability and competitiveness versus modern transformer backbones. Performance is weaker on synthetic CAMO images (the paper attributes this to distorted part–whole relations), suggesting sensitivity to certain data distributions. Code is stated as

	includes an analysis splitting CAMO into real and synthetic subsets to explain failure modes.	
[35]	Strongly motivated shift from generic "context" to scene-level positional relationship modeling, implemented concretely via GCN with self-attention and reverse self-attention adjacency (PAM). The fusion design is also well-justified: IRM explicitly accounts for semantic correlation and detail discontinuity, and ablations show consistent gains from PAM, FIM, and IRM. The paper additionally demonstrates transfer to related segmentation tasks (transparent object segmentation and polyp segmentation), supporting methodological generality.	The paper notes that performance on NC4K can be slightly behind the top-ranked methods and reports failure cases under severe dim backgrounds, heavy occlusion, and weak surrounding semantics. It also explicitly acknowledges that the current design prioritizes accuracy and does not optimize computational efficiency. Data are "available on request," and public code availability is not clearly stated, which can limit reproducibility.
[36]	Clear paradigm shift from deterministic decoding to iterative denoising, with a well-motivated conditioning mechanism (FF + IAM) that is validated by ablations. Evaluated on four standard benchmarks with comparisons against 11 strong baselines and includes qualitative evidence emphasizing fine texture recovery. Code release is explicitly stated with a public repository link.	Diffusion inference is inherently more expensive than single-pass models due to iterative sampling (T-step formulation), which can limit real-time deployment. Gains depend on diffusion configuration (steps, schedule) and conditioning quality, which increases sensitivity to implementation choices. The evaluation is benchmark-centric and does not include a dedicated cross-domain robustness protocol beyond the standard dataset suite.
[37]	The method is conceptually coherent and easy to justify: each stage targets a specific camouflage difficulty, and the design choices are backed by extensive ablations (dual-domain input, RSM window size, uncertainty fusion variants, CPM iteration count, and supervision weighting). The RSM offers a practical alternative to dense CRF connectivity by using block-level, window-restricted interactions, and UEM adds an explicit mechanism for ambiguous boundaries. The paper reports both accuracy and runtime/parameter comparisons, improving auditability of improvements.	Core evaluation remains within standard benchmarks; although the method discusses challenging cases and presents auxiliary polyp segmentation transfer, it does not formalize a dedicated cross-domain generalization protocol for camouflage beyond the usual dataset suite. The three-stage design (dual-domain encoding plus scanning and iterative recalibration) increases architectural complexity, and performance may be sensitive to implementation choices (window size, number of scan modules, CPM iterations). Data availability is stated as "on request," which can limit full reproducibility.
[38]	Clear, defensible decomposition of the task into boundary localization and position reasoning before context fusion; introduces a concrete attention mechanism (Retrieve Attention) and validates it against alternative boundary modules; includes structured ablations for EEM/OPRM and dilation-channel design; evaluates multiple backbones and provides a generalization experiment on polyp segmentation datasets.	Not designed for efficiency: the best setting reports 105.2M parameters and 65.7G FLOPs, explicitly prioritizing accuracy over lightweight deployment; evaluation is centered on three standard datasets without a dedicated cross-domain robustness protocol; data are stated as "available on request," and public code availability is not clearly specified in the manuscript.

Typology-Driven Synthesis and Mechanism-Based Analysis

Looking across the 38 studies in Table 1, one thing is hard to miss: most papers still validate primarily through the usual benchmark circuit (CAMO, COD10K, NC4K, CHAMELEON). Method choices also cluster rather than scatter. Transformer-enabled designs show up often (20/38, 52.6%), multi-scale or staged refinement is even more common (31/38, 81.6%), and explicitly boundary/edge-aware modeling appears in a smaller—but still meaningful—slice (10/38, 26.3%). What is less consistent is how results are reported: code release and efficiency indicators are each available for only 15/38 studies (39.5%), and explicit robustness or domain-shift checks remain relatively rare (7/38, 18.4%).

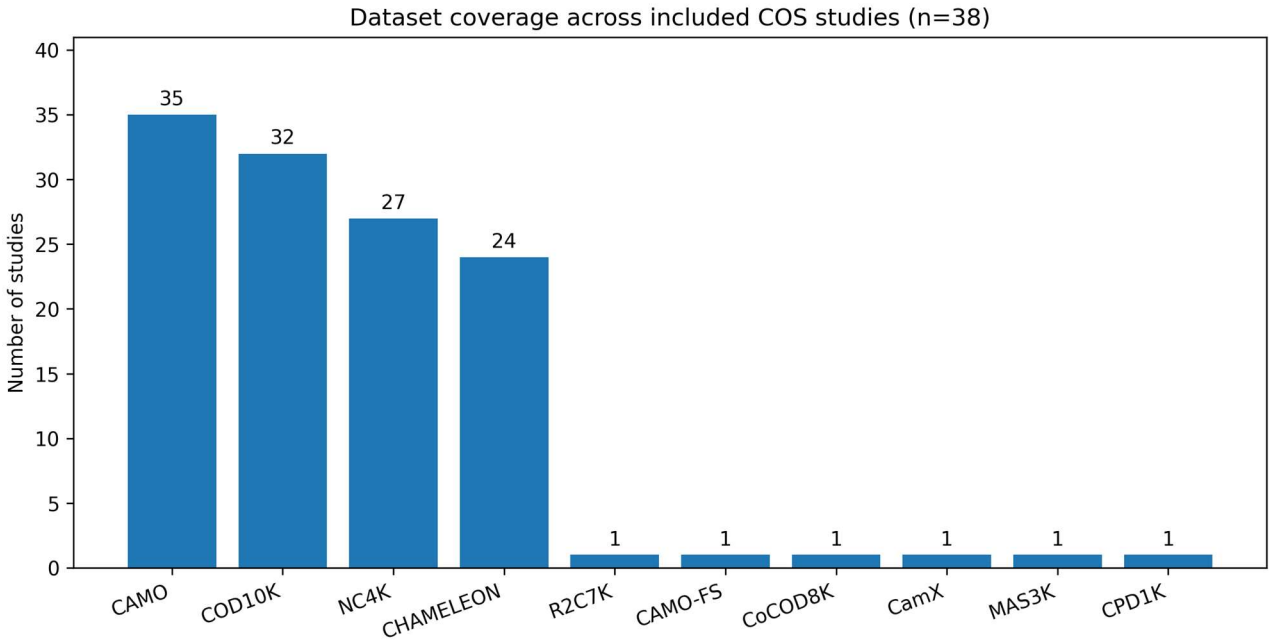

Figure 3- Dataset coverage across the included COS studies (n=38). Most evaluations focus on the standard benchmark circuit (CAMO, COD10K, NC4K, and CHAMELEON), while task-specific

Given that many methods combine several ideas, the most informative way to read Table 1 is not "which architecture family won," but "what mechanism each method leans on to control errors." In practice, COS papers keep running into the same two headaches: boundaries are weak (so predictions leak), and foreground/background are too similar (so distractors cause false positives or misses). The typology below follows that logic. It is intentionally multi-label—because most papers are multi-idea—but each family is defined by the main error-control mechanism that the method seems to be

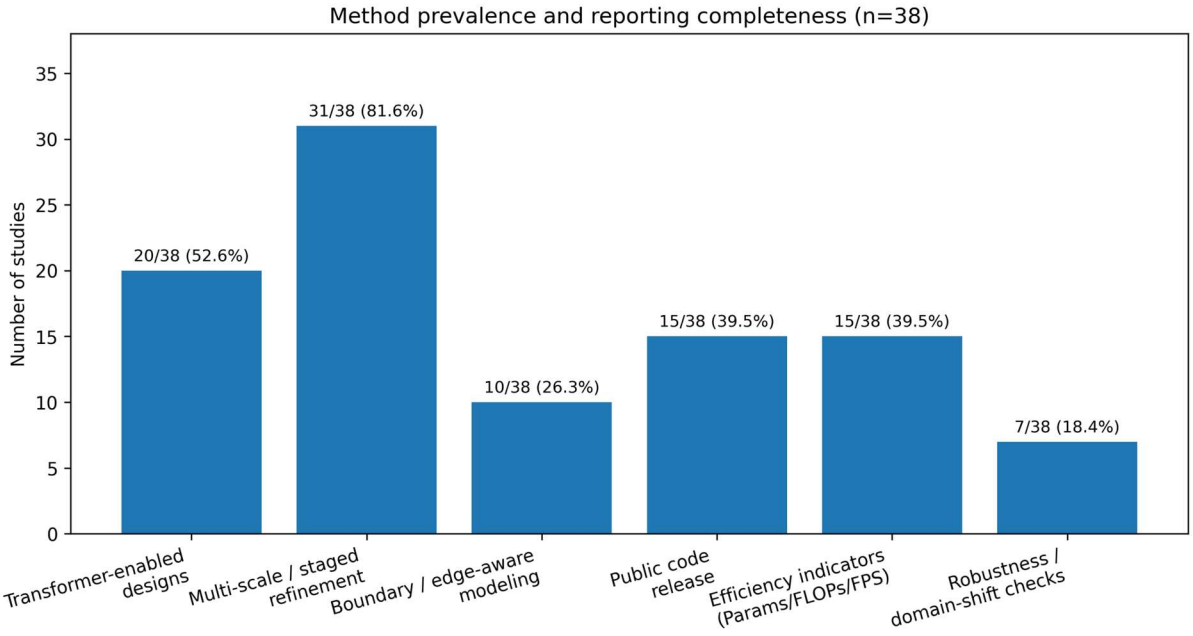

Figure 4- Prevalence of core design patterns and reporting completeness (n=38)

built around.

Boundary and structure as a first-class objective

A recurring group of strong entries in Table 1 clearly treats boundary control as "the core problem," not a finishing touch. You see this in explicit edge semantics and edge-guided fusion ([1], [2]), and also in the many "coarse prior first, boundary/detail later" designs ([22], [38]). The shared intuition is simple: if the contour is wrong, everything downstream looks wrong—especially in cluttered scenes where background textures imitate the object.

What seems to matter here is not just having an edge branch, but whether it is used in a way that actually suppresses distractor edges. The more convincing papers are usually the ones that connect boundary signals to refinement (rather than predicting edges in isolation), or pair them with coarse localization so the model is less likely to chase background lines ([2], [22], [38]). That pairing also lines up with the kind of improvements these papers emphasize: fewer leakages, cleaner structures, and more stable masks.

Multi-scale / staged refinement: the default design pattern

If there is a "default recipe" in Table 1, it is not a backbone—it is coarse-to-fine reasoning. Mixed-scale zooming ([3], [33]), global-to-local cascades ([10], [31], [38]), stepwise refinement that repeatedly repairs incomplete masks ([4]), multi-view inspection and fusion ([28]), and even mirror-stream reasoning ([18]) all fit the same bigger picture: camouflage cues are sparse, fragile, and often visible only under the right scale or view.

What is nice about the stronger papers in this family is that they do not just say "we used more stages." They tend to separate roles: early parts handle "where is it, roughly," and later parts handle "make it complete and clean." That makes the improvement story easier to believe, and it also makes ablations more meaningful (you can actually see what each stage buys you) ([3], [4], [10], [28], [38]).

One caveat that Table 1 hints at (and Table 2 usually reinforces) is that some gains can be tightly coupled to test-time recipes: multi-scale inference, repeated refinement, aggregation tricks. When papers are transparent about those settings, it is fine; when they are not, cross-paper comparisons become less reliable.

Global context and priors: transformers help, but only when they do something specific

Transformers are everywhere in this set, but Table 1 does not really support the idea that "transformer = better" on its own. The methods that read as more solid are usually those that spell out how global context is used to guide refinement. You can see a few clear patterns here:

Explicit CNN–Transformer interaction where transformers provide global positioning and CNN features preserve texture/boundary fidelity ([5], and efficiency-minded hybrids like [19], [30]).

Prior generation + prior guidance done deliberately rather than implicitly ([20]).

Cascades where a transformer-heavy stage provides a reliable coarse map, and a lighter refinement stage focuses on detail and boundaries ([10]).

So, the "lesson" is not that transformers dominate, but that guided use of global priors (interaction, staged guidance, or explicit priors) tends to show up in papers with clearer improvement narratives.

Ambiguity-aware iterative estimation: unfolding and diffusion

A smaller—but conceptually important—cluster treats camouflage as an ambiguity problem where one-shot masks are often overconfident. Here the table includes deep unfolding with reversible correction ([12]) and multiple diffusion-style pipelines ([6], [36]), including two-stage hybrids that use diffusion to build a prior and a second branch to sharpen structure ([8]). There is also diffusion coupled with depth as guidance ([7]).

These methods often look strong on accuracy metrics, and the motivation makes sense: if the model is allowed to "revise" the mask through iterative correction, it has a built-in way to clean up uncertain regions. At the same time, this is also where reporting details become crucial. Sampling steps, ensembles, schedules—small changes can matter. Without full transparency, it becomes hard to tell whether we are comparing methods or comparing inference settings.

Extra cues and cross-task constraints

Table 1 also includes methods that bring in additional signals to break the deadlock of low contrast: depth-assisted learning ([11]) and diffusion-coupled depth guidance ([7]) are the obvious cases; frequency cues with region-level coherence and uncertainty gating are another ([37]). There is also cross-task modeling that explicitly leverages conflicts/consistencies between SOD and COD signals ([21]).

These approaches are often practical: extra cues can make ambiguous regions less ambiguous. The trade-off is that performance may depend on the quality (and cost) of the auxiliary pipeline. So, the fairest interpretation is usually: "strong in practice, but carries extra dependencies," unless the paper carefully reports sensitivity to cue quality.

Training-regime contributions: distillation, adversarial augmentation, and contrastive separation

Finally, several entries in Table 1 improve COS not by a flashy decoder block, but by changing how representations are learned:

- Distillation explicitly used to handle imbalance and transfer strong performance to lighter students ([9]).
- Adversarial sample generation that makes camouflage harder during training to improve generalizability ([14]).
- Contrastive learning that forces separation between camouflaged objects, background, and salient non-camouflaged objects ([27]).

These papers matter for the synthesis because they show that "what works" is sometimes a learning mechanism rather than an architecture trick—especially when the goal is robustness or deployability.

Task variants that broaden the evidence

Beyond standard single-image binary COS, Table 1 includes referring discovery ([15]) to reduce false positives when the object is absent, collaborative/group detection with a new dataset ([24]), few-shot instance segmentation ([23]), instance-level camouflage segmentation in a Mask2Former-style setting ([26]), underwater camouflage-like segmentation under style shifts ([32]), and generalization-oriented evaluation via CamX ([25]).

Even when these are not directly "rank-comparable" to the base setting, they matter because they reveal real failure cases that standard benchmarks do not always stress (no-object scenes, group consensus, label scarcity, domain degradation).

What the table supports and what it does not

If Table 1 supports one high-level message, it is this: the most repeatable gains tend to come from

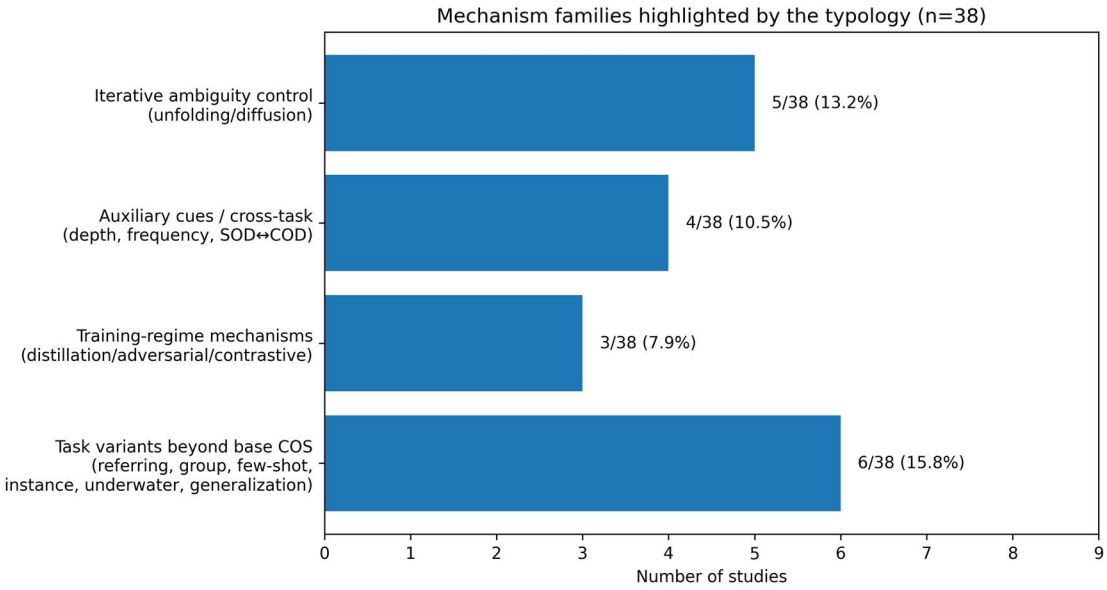

Figure 5- Iterative ambiguity-control approaches (unfolding/diffusion), auxiliary cues (e.g., depth/frequency/cross-task constraints), training-regime mechanisms (distillation/adversarial/contrastive), and task variants beyond the base COS setup occur less frequency

combining (i) reliable coarse localization (often via global priors), (ii) explicit boundary/structure control, and (iii) multi-scale or staged refinement. That combination keeps appearing across very different design choices, and it also explains why “just swapping a backbone” rarely sounds like a complete story.

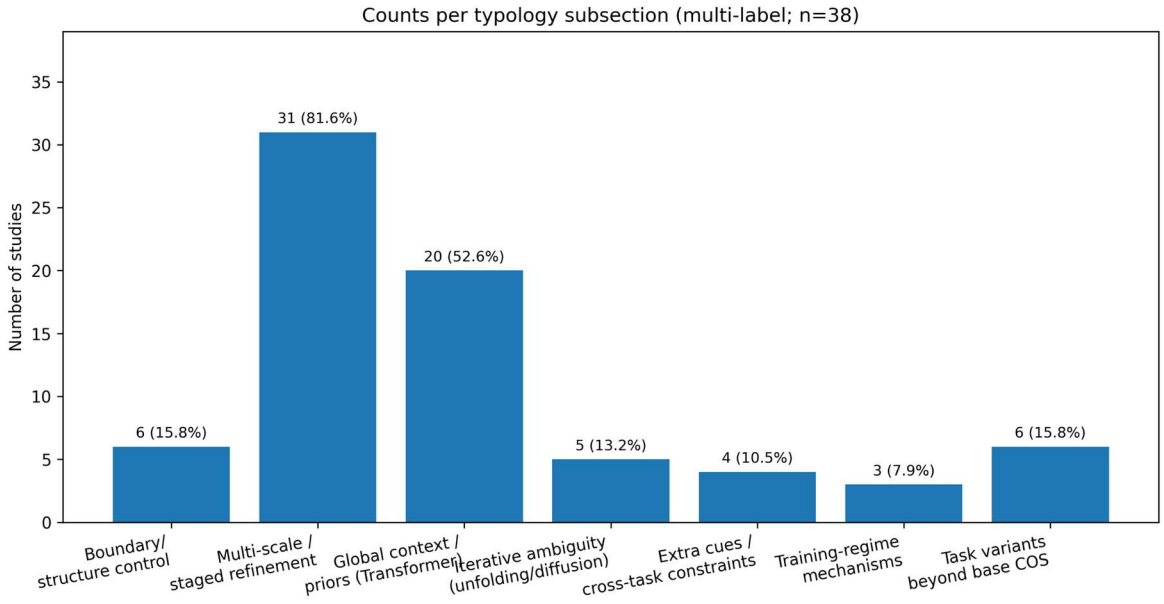

Figure 6- The typology labels are not mutually exclusive; a single study may contribute to multiple mechanism categories. Counts are reported to show which mechanisms are most frequently emphasized across the included evidence base (Table 1; Section 3.3).

What Table 1 does not support well is strict cross-paper ranking. Protocol differences are too large (resolution, inference tricks, diffusion steps, auxiliary pipelines), and reporting is too inconsistent (code/efficiency/robustness are missing often enough) to treat a leaderboard view as fully reliable. For that reason, the most defensible contribution of the review is a mechanism-level synthesis: which mechanisms repeatedly align with gains, and which trade-offs (compute cost, inference sensitivity, auxiliary dependencies) keep showing up.

4 Disscussion

Over the period covered in this review, progress in camouflaged object segmentation appears to be driven less by simply swapping backbone networks and more by adding mechanisms that match the problem's specific failure modes. In particular, the literature increasingly converges on the idea that meaningful gains come when methods address two difficulties at the same time: (i) weak or ambiguous boundaries that cause background leakage, and (ii) extreme foreground–background similarity that makes distractors hard to reject. In that sense, the strongest line of progress is fairly consistent with what Table 1 reveals: boundary/structure control, staged or multi-scale refinement for recovering subtle cues, and global context modeling that does not erase local detail tend to work best when they are designed to interact rather than operate in isolation. More recent iterative formulations—most notably diffusion-style refinement and unfolding-style correction—also fit naturally into this trajectory, because they treat camouflage as an ambiguity problem where single-shot masks are often overconfident. At the same time, these iterative approaches highlight a central trade-off: as ambiguity handling improves, inference cost and configuration sensitivity usually

increase, which makes cross-paper comparison harder unless evaluation settings are reported and standardized.

At the evidence level, the review also makes clear why strict cross-paper ranking is difficult to justify at present. Most studies remain benchmark-centric, with evaluation concentrated on a small set of widely used datasets, and explicit domain-shift or robustness testing is still limited. Comparability is further weakened by protocol variability that can materially affect reported numbers: backbone capacity, input resolution, multi-scale testing, ensembling or sampling strategies (especially for diffusion-based methods), auxiliary supervision pipelines, and even differences in the task definition (standard COS versus variants such as referring, collaborative, few-shot, or instance-level settings). As a result, reported improvements should be interpreted as holding under each paper's stated experimental configuration, rather than as universally superior behavior across settings. The lack of routine uncertainty quantification (e.g., confidence intervals or calibrated uncertainty reporting) reinforces the view that the most defensible synthesis at this stage is mechanism-level: identifying patterns of qualitative agreement across method families, rather than claiming a definitive global ranking.

Reproducibility and transparency further limit evidential strength. While many papers provide strong architectural descriptions and ablation studies, code release and full reporting of training and inference recipes remain inconsistent, and deployment-oriented reporting (e.g., FPS/FLOPs and hardware specifications) varies widely. This raises a practical concern: some apparent progress may be partly explained by unreported implementation details or favorable inference recipes rather than a clean methodological advantage. For COS research in particular—where test-time choices can noticeably move the needle—protocol clarity and reproducibility should be treated as first-order quality factors, alongside accuracy.

The review process itself also has limitations that should be acknowledged. Screening and extraction were performed by a single reviewer, which can increase the risk of missed studies and subjective classification, especially in a literature where COS and COD terminology is sometimes used loosely. To mitigate this, ambiguous records were conservatively retained for full-text screening, and a random subset was re-screened after a two-week interval to improve internal consistency. Nevertheless, some selection bias cannot be fully ruled out. In addition, assigning a primary methodological type inevitably simplifies multi-idea methods, even when a multi-label typology (as used in Section 3.3) helps reduce that distortion.

In terms of practical implications, the most defensible takeaway is that COS systems benefit most from designs that explicitly control boundaries and suppress distractors, particularly when paired with multi-scale or staged reasoning that can exploit subtle cues not visible at a single receptive field (Table 1). In the near term, efficiency-oriented designs and compression-focused strategies are likely the most straightforward path toward deployment. In contrast, diffusion, multi-view, and other iterative refinement approaches should be viewed as accuracy-forward directions whose computational cost and sensitivity to inference settings must be managed explicitly. Looking ahead, three directions appear especially important: (i) standardized reporting of training and inference settings (resolution, augmentation, test-time tricks, sampling steps, and cost), (ii) broader robustness testing to better support generalization claims, and (iii) uncertainty-aware evaluation that reflects the inherent ambiguity of camouflage and enables more informed assessment beyond point estimates.

5 Conclusion

This systematic literature review aggregates the research on camouflaged object segmentation (COS) between 2020 and 2026, employing a method-centric taxonomy and an appraisal-oriented approach for 38 eligible studies. Besides presenting an overview of the performance on standard benchmarks, this review aims to provide an integration of common modeling mechanisms and an identification of areas where empirical progress appears most consistent across datasets and method families.

Strengths and Practical Value. The most notable advantage of this review is that it attempts to organize the disparate research on COS into an organized, method-driven framework that can be used as an effective reference guide. By employing this taxonomy and standardizing the review process, this literature review allows the reader to (i) recognize the most popular design trends behind reported performance gains (e.g., boundary/structure-aware refinement, multi-scale/staged reasoning, transformer-enabled global context information combined with local detail recovery, and uncertainty-aware iterative refinement), (ii) make comparisons between different studies by making dataset/protocol choices explicit, and (iii) interpret performance gains in the context of reporting completeness (e.g., code availability, efficiency indicators, and robustness tests).

Limitations of this review. There are a number of limitations to the current synthesis, which should be borne in mind when drawing conclusions from the results. In particular, the review process relied only on a single reviewer for the screening and extraction process, which may have introduced a number of potential biases. In addition, it was not possible to retrieve the full text for a number of the papers, and the current evidence base is heterogeneous with regard to backbone architectures, resolutions, test time settings (such as multi-scale testing or the use of ensembling or diffusion sampling), and even task variations. As such, the current results should be treated as being most reliable when they demonstrate qualitative consensus across method families rather than as definitive statements about the absolute best-performing model.

Implications and Future Directions. The findings suggest that basic requirements for making actual progress on COS are no longer just related to accuracy but also relate to how well the entire approach is communicated and how costs are taken into consideration. With this background, future directions and how they can make COS accuracy and progress even sharper include: (i) Designing standardized reporting schemes for training and test processes, e.g., resolution levels, augmentation strategies, test-time tricks if used, number of sampling steps, and overall computational costs involved; (ii) Testing robustness beyond benchmark accuracy to domain shifts and carefully crafted test scenarios these would provide a much better perspective on how well these approaches generalize and hold up to malice and mischief; and (iii) Including uncertainty-aware evaluation and quantifying failure probability part of understanding any camouflage scene inherently involves quantifying the probability of ambiguity and failure.

References

[1] Y. Sun, S. Wang, C. Chen, and T.-Z. Xiang, "Boundary-guided camouflaged object detection," presented at the Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, 2022, 2022. [Online]. Available: *https://doi.org/10.24963/ijcai.2022/186*.

[2] B. Fan, K. Cong, and W. Zou, "Dual Attention and Edge Refinement Network for Camouflaged Object Detection," in *2023 8th International Conference on Image, Vision and Computing (ICIVC)*, 27-29 July 2023 2023, pp. 60-65, doi: 10.1109/ICIVC58118.2023.10270622.

[3] Y. Pang, X. Zhao, T. Z. Xiang, L. Zhang, and H. Lu, "ZoomNeXt: A Unified Collaborative Pyramid Network for Camouflaged Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 46, no. 12, pp. 9205-9220, 2024, doi: 10.1109/TPAMI.2024.3417329.

[4] X. Yan, M. Sun, Y. Han, and Z. Wang, "Camouflaged Object Segmentation Based on Matching–Recognition–Refinement Network," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 35, no. 11, pp. 15993-16007, 2024, doi: 10.1109/TNNLS.2023.3291595.

[5] F. Wu, X. Li, Y. Zhang, and K. Hu, "TransCoop: Cooperation of Transformers and CNNs for Camouflaged Object Segmentation," in *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 18-22 July 2022 2022, pp. 1-6, doi: 10.1109/ICME52920.2022.9859746.

[6] Z. Chen, K. Sun, and X. Lin, "CamoDiffusion: Camouflaged object detection via conditional diffusion models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, no. 2, pp. 1272-1280.

[7] X. Jiang *et al.*, "BiDiCOS: Camouflaged object segmentation via bilateral diffusion model," *Expert Systems with Applications,* vol. 255, p. 124747, 2024/12/01/ 2024, doi: *https://doi.org/10.1016/j.eswa.2024.124747*.

[8] H. Zhang and S. Jiang, "SDNet: A two-stage segmentation-diffusion network for camouflaged object detection," *Applied Soft Computing,* Article vol. 188, 2026, Art no. 114390, doi: 10.1016/j.asoc.2025.114390.

[9] B. Cai, H. Li, Y. Yang, and J. Yan, "CFF-KDNet: Cross-scale feature fusion network with knowledge distillation for camouflaged object detection," *Expert Systems with Applications,* vol. 299, p. 130209, 2026/03/01/ 2026, doi: *https://doi.org/10.1016/j.eswa.2025.130209*.

[10] Y. Zhao *et al.*, "HRTNet: Holistic registration theory-inspired network for camouflaged object detection," *Neurocomputing,* vol. 671, p. 132674, 2026/03/28/ 2026, doi: *https://doi.org/10.1016/j.neucom.2026.132674*.

[11] X. Li, X. Yu, and P. Chen, "DSANet: Deep surrounding-aware network for camouflaged object detection via cross-refinement mirror strategy," *Expert Systems with Applications,* vol. 298, p. 129605, 2026/03/01/ 2026, doi: *https://doi.org/10.1016/j.eswa.2025.129605*.

[12] C. He *et al.*, "RUN: Reversible Unfolding Network for Concealed Object Segmentation," presented at the Proceedings of Machine Learning Research, Proceedings of Machine Learning Research, 2025. [Online]. Available: *https://proceedings.mlr.press/v267/he25w.html*.

[13] J. Du *et al.*, "Shift the Lens: Environment-Aware Unsupervised Camouflaged Object Detection," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10-17 June 2025 2025, pp. 19271-19282, doi: 10.1109/CVPR52734.2025.01795.

[14] C. He *et al.*, "Strategic Preys Make Acute Predators: Enhancing Camouflaged Object Detectors by Generating Camouflaged Objects," presented at the International Conference on Learning Representations, 2024, 2024. [Online]. Available: *https://proceedings.iclr.cc/paper_files/paper/2024/file/9e1fe089a2b3fee44a4043fa6830c00f-Paper-Conference.pdf*.

[15] A. Gupta, K. R. Jerripothula, and T. Tillo, "CIRCOD: Co-Saliency Inspired Referring Camouflaged Object Discovery," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 26 Feb.-6 March 2025 2025, pp. 8313-8323, doi: 10.1109/WACV61041.2025.00806.

[16] J. He, B. Liu, and H. Chen, "HDPNet: Hourglass Vision Transformer with Dual-Path Feature Pyramid for Camouflaged Object Detection," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 26 Feb.-6 March 2025 2025, pp. 8638-8647, doi: 10.1109/WACV61041.2025.00837.

[17] H. Mei, G. P. Ji, Z. Wei, X. Yang, X. Wei, and D. P. Fan, "Camouflaged Object Segmentation with Distraction Mining," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 20-25 June 2021 2021, pp. 8768-8777, doi: 10.1109/CVPR46437.2021.00866.

[18] J. Yan, T.-N. Le, K.-D. Nguyen, M.-T. Tran, T.-T. Do, and T. V. Nguyen, "Mirrornet: Bio-inspired camouflaged object segmentation," *IEEE access,* vol. 9, pp. 43290-43300, 2021.

[19] X. Hu, X. Zhang, F. Wang, J. Sun, and F. Sun, "Efficient Camouflaged Object Detection Network Based on Global Localization Perception and Local Guidance Refinement," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 34, no. 7, pp. 5452-5465, 2024, doi: 10.1109/TCSVT.2023.3349209.

[20] R. Wang *et al.*, "Camouflaged object segmentation with prior via two-stage training," *Computer Vision and Image Understanding,* vol. 246, p. 104061, 2024/09/01/ 2024, doi: *https://doi.org/10.1016/j.cviu.2024.104061*.

[21] Y. Yang and Q. Zhang, "Finding Camouflaged Objects Along the Camouflage Mechanisms," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 34, no. 4, pp. 2346-2360, 2024, doi: 10.1109/TCSVT.2023.3308964.

[22] C. Wang, Y. Li, G. Wei, X. Hou, and X. Sun, "Robust Localization-Guided Dual-Branch Network for Camouflaged Object Segmentation," *Electronics,* vol. 13, no. 5, p. 821, 2024. [Online]. Available: *https://www.mdpi.com/2079-9292/13/5/821*.

[23] T. D. Nguyen *et al.*, "The Art of Camouflage: Few-Shot Learning for Animal Detection and Segmentation," *IEEE Access,* vol. 12, pp. 103488-103503, 2024, doi: 10.1109/ACCESS.2024.3432873.

[24] C. Zhang, H. Bi, T. Z. Xiang, R. Wu, J. Tong, and X. Wang, "Collaborative Camouflaged Object Detection: A Large-Scale Dataset and Benchmark," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 35, no. 12, pp. 18470-18484, 2024, doi: 10.1109/TNNLS.2023.3317091.

[25] L. Xu, X. You, F. Jia, and K. Liu, "BiCOD: A Camouflaged Object Detection Method Directed by Cognitive Attention," *IEEE Sensors Journal,* vol. 24, no. 4, pp. 4711-4721, 2024, doi: 10.1109/JSEN.2023.3343917.

[26] T.-H. Phung and H.-H. Shuai, "Revealing Hidden Context in Camouflage Instance Segmentation," in *Computer Vision – ACCV 2024*, Singapore, M. Cho, I. Laptev, D. Tran, A. Yao, and H. Zha, Eds., 2025// 2025: Springer Nature Singapore, pp. 3-20.

[27] X. Jiang *et al.*, "Camouflaged Object Segmentation Based on Joint Salient Object for Contrastive Learning," *IEEE Transactions on Instrumentation and Measurement,* vol. 72, pp. 1-16, 2023, doi: 10.1109/TIM.2023.3306520.

[28] D. Zheng, X. Zheng, L. T. Yang, Y. Gao, C. Zhu, and Y. Ruan, "Mffn: Multi-view feature fusion network for camouflaged object detection," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 6232-6242.

[29] F. Wu, J. Yin, X. Li, J. Wu, D. Jin, and J. Yang, "CoNet: A Consistency-Oriented Network for Camouflaged Object Segmentation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 35, no. 1, pp. 287-299, 2025, doi: 10.1109/TCSVT.2024.3462465.

[30] Y. Yuan, K. Zhang, and J. Zhang, *Camouflaged Object Detection Through Feature Selection and Enhancement*. 2024, pp. 452-457.

[31] K. Wang, H. Bi, Y. Zhang, C. Zhang, Z. Liu, and S. Zheng, "D2C-Net: A Dual-branch, Dual-guidance and Cross-refine Network for Camouflaged Object Detection," (in English), *IEEE Transactions on Industrial Electronics,* vol. 69, no. 5, pp. 5364-5374, 2022-05 2022, doi: 10.1109/tie.2021.3078379.

[32] R. Chen, Z. Fu, Y. Huang, E. Cheng, and X. Ding, "A Robust Object Segmentation Network for UnderWater Scenes," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 23-27 May 2022 2022, pp. 2629-2633, doi: 10.1109/ICASSP43922.2022.9746176.

[33] X. Li, L. Li, S. Jiang, M. Yang, and L. Qi, "Camouflaged Object Detection with Discriminative Information Attention and Cross-level Feature Fusion," in *2022 7th International Conference on Image, Vision and Computing (ICIVC)*, 26-28 July 2022 2022, pp. 248-255, doi: 10.1109/ICIVC55077.2022.9886094.

[34] Y. Liu, D. Zhang, Q. Zhang, and J. Han, "Integrating Part-Object Relationship and Contrast for Camouflaged Object Detection," *IEEE Transactions on Information Forensics and Security,* vol. 16, pp. 5154-5166, 2021, doi: 10.1109/TIFS.2021.3124734.

[35] J. Zhang, G. Yang, X. Dai, and P. Yang, "SPANet: Spatial perceptual activation network for camouflaged object detection," *IET Computer Vision,* vol. 18, pp. 1300-1312, 09/18 2024, doi: 10.1049/cvi2.12310.

[36] Z. Chen, R. Gao, T.-Z. Xiang, and F. Lin, "Diffusion Model for Camouflaged Object Detection," 2023.

[37] Y. Liu, J. Zhang, Y. Wang, and J. Jin, "CamoSSR: A unified framework with energy-based scanning and cognitive reasoning for camouflaged object detection," *Neurocomputing,* vol. 663, p. 131977, 2026/01/28/ 2026, doi: *https://doi.org/10.1016/j.neucom.2025.131977*.

[38] S. Jiang, H. Zhang, and C. Ding, "Dual-optimized two-stage Camouflaged Object Detection," *Journal of Visual Communication and Image Representation,* vol. 114, p. 104631, 2026/01/01/ 2026, doi: *https://doi.org/10.1016/j.jvcir.2025.104631*.