

# Refining Context-Entangled Content Segmentation via Curriculum Selection and Anti-Curriculum Promotion

Chunming He<sup>1</sup> Rihan Zhang<sup>1</sup> Fengyang Xiao<sup>1</sup> Dingming Zhang<sup>1</sup> Zhiwen Cao<sup>2</sup> Sina Farsiu<sup>1</sup>

## Abstract

Biological learning proceeds from easy to difficult tasks, gradually reinforcing perception and robustness. Inspired by this principle, we address Context-Entangled Content Segmentation (CECS), a challenging setting where objects share intrinsic visual patterns with their surroundings, as in camouflaged object detection. Conventional segmentation networks predominantly rely on architectural enhancements but often ignore the learning dynamics that govern robustness under entangled data distributions. We introduce CurriSeg, a dual-phase learning framework that unifies curriculum and anti-curriculum principles to improve representation reliability. In the Curriculum Selection phase, CurriSeg dynamically selects training data based on the temporal statistics of sample losses, distinguishing hard-but-informative samples from noisy or ambiguous ones, thus enabling stable capability enhancement. In the Anti-Curriculum Promotion phase, we design Spectral-Blindness Fine-Tuning, which suppresses high-frequency components to enforce dependence on low-frequency structural and contextual cues and thus strengthens generalization. Extensive experiments demonstrate that CurriSeg achieves consistent improvements across diverse CECS benchmarks without adding parameters or increasing total training time, offering a principled view of how progression and challenge interplay to foster robust and context-aware segmentation. The code is available at <https://github.com/ChunmingHe/CurriSeg>.

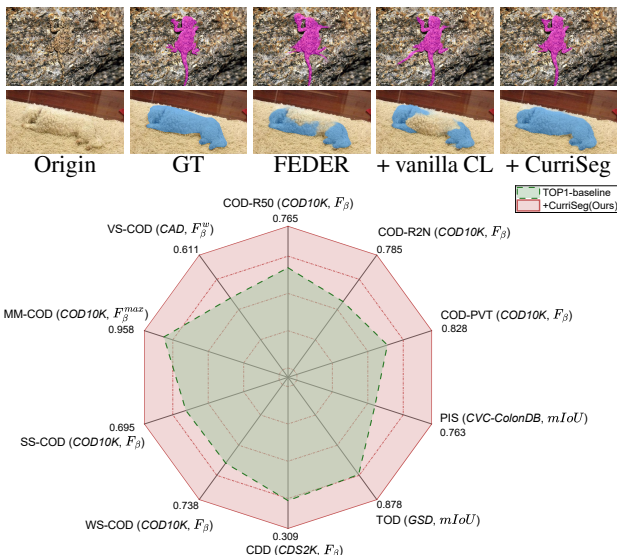


Figure 1. Performance on CECS data. **Top**: CL is short for curriculum learning (Bengio et al., 2009). Concealed object masks are highlighted in purple and blue, overlaid on the original data for visual clarity. **Bottom**: “Task (Dataset, Metric)”. The radar chart shows that training the baseline under our CurriSeg brings performance gains than the standard training manner, where TOP-1 is the composite baselines with the top metric scores per task.

## 1. Introduction

Context-Entangled Content Segmentation (CECS) aims to segment those extremely concealed objects that share intrinsic visual patterns with their surroundings (Fan et al., 2020; Zhao et al., 2024; He et al., 2024b; Lu et al., 2024; Xiao et al., 2024; He et al., 2026), with applications in medical analysis (Tajbakhsh et al., 2015; Hu et al., 2026; He et al., 2025b), autonomous perception (Mei et al., 2020; Hu et al., 2025; Gao et al., 2025; Zheng et al., 2025), etc.

Unlike generic segmentation, CECS suffers from extreme feature ambiguity. Existing methods mainly focus on architectural engineering, such as feature aggregation (Pang et al., 2022; Fan et al., 2020) complex decoders (He et al., 2025e;c), or prior injection (Xiao et al., 2026b;a), yet overlook a fundamental bottleneck: the learning dynamics when facing entangled data distributions.

A natural strategy to handle such complexity is Curriculum Learning (CL) (Bengio et al., 2009), learning from easy to

<sup>1</sup>Duke University <sup>2</sup>Adobe. Correspondence to: Fengyang Xiao <fengyang.xiao@duke.edu>, Sina Farsiu <sina.farsiu@duke.edu>.

hard, mimicking biological cognition. However, we argue that standard CL is detrimental in CECS scenarios. In camouflaged scenes, “easy” samples often exhibit high spurious correlations (*e.g.*, distinct background textures that unintentionally reveal the object). Training on these first drives the model into lazy regimes that rely on high-frequency texture bias rather than semantic structure, leading to poor generalization on hard samples, as validated in Fig. 1.

To address this, we introduce CurriSeg, a novel framework that orchestrates training through two phases: Robust Curriculum Selection and Anti-Curriculum Promotion. Inspired by biological mastery, where predators first acquire fundamental skills then adapt to challenging environments, CurriSeg stabilizes optimization before deliberately “blinding” the model to encourage deeper feature extraction.

In the Curriculum Selection phase, we address the noise–ambiguity dilemma inherent in CECS. Instead of relying on static difficulty measures, we monitor temporal behavior of samples. By tracking mean and variance of sample losses, CurriSeg distinguishes hard-but-informative samples from noisy or ambiguous ones. We also incorporate pixel-level uncertainty estimation to prevent uncertain regions from dominating early gradients. Combined with a warm-up curriculum strategy, this yields a joint image- and pixel-level curriculum that is stable and noise-resistant.

After the model reaches a stable regime, we introduce Anti-Curriculum Promotion to enhance robustness on hard samples. We propose Spectral-Blindness Fine-Tuning (SBFT), which deliberately attenuates high-frequency input components. By temporarily suppressing texture cues, SBFT forces the model to rely on low-frequency structural patterns and contextual semantics, promoting extraction of intrinsic, task-relevant signals. This explicitly counteracts the tendency to exploit superficial texture shortcuts and encourages discovery of complementary, subtler cues, improving robustness where texture-based heuristics fail.

Together, these two phases establish a principled paradigm that stabilizes optimization then intentionally increases difficulty to overcome representational bottlenecks. CurriSeg combines curriculum and anti-curriculum learning within a unified framework, offering a new perspective on training for context-entangled visual tasks.

Our contributions are summarized as follows:

- (1) We propose CurriSeg, the first CL-based CECS paradigm. By following a “stabilize-then-perturb” trajectory, it effectively breaks the learning bottleneck caused by feature ambiguity in concealed scenes.
- (2) We design robust curriculum selection that uses temporal loss statistics and uncertainty-aware pixel masking to construct a stable, noise-resistant learning trajectory.

- (3) We introduce an anti-curriculum promotion phase, SBFT, which attenuates high-frequency cues and encourages the model to exploit complementary, subtle signals, thereby improving robustness on challenging camouflaged cases.

- (4) We validate CurriSeg on multiple CECS benchmarks and backbones, showing consistent performance gains. CurriSeg adds no extra parameters and maintains training cost comparable to standard training schedules.

## 2. Related Works

**Context-entangled content segmentation.** Existing CECS methods mainly focus on structural designs: encoder-decoder backbones with multi-scale aggregation (Fan et al., 2020; 2021a; Li & Fu, 2025), attention-based refinement (Pang et al., 2022; 2024), edge- or uncertainty-aware branches (He et al., 2023a; 2025d), transformer-style context modeling (Sun et al., 2024; Zhang et al., 2025), and frequency-aware modules (He et al., 2025f; Shen et al., 2025). They enhance feature representation at the network level, yet training typically follows a standard supervised routine with shuffled mini-batches. The roles of sample difficulty, label ambiguity, and texture shortcuts in shaping CECS learning dynamics remain underexplored. Our work instead organizes the training procedure via curriculum and anti-curriculum mechanisms, orthogonal to existing architectural advances.

**Curriculum learning.** CL organizes training samples from easy to hard to improve generalization (Bengio et al., 2009; Soviany et al., 2022). Variants include self-paced learning (Kumar et al., 2010; Tullis & Benjamin, 2011), teacher–student curricula (Matiisen et al., 2019; Saglietti et al., 2022), and task-specific schemes via loss-based ranking (Pentina et al., 2015) or hardness-aware sampling (Soviany, 2020). However, CL assumes easy samples provide reliable supervision and difficulty correlates with learning utility. In CECS, these assumptions break down: “easy” samples may contain spurious textures promoting shortcuts, while hard samples can be both informative and ambiguous. Moreover, most CL methods overlook the pixel-level nature of segmentation with spatially localized uncertainty. CurriSeg departs from CL by (i) using temporal loss statistics to distinguish hard-but-informative samples from noisy cases, and (ii) adding an anti-curriculum phase that increases difficulty via spectral manipulation, mitigating texture bias.

## 3. Methodology

CurriSeg is a dual-phase learning framework designed to address the feature ambiguity inherent in CECS. As illustrated in Fig. 2 and Algorithm 1, the framework follows a “stabilize-then-perturb” trajectory comprising two sequential phases: (1) Robust Curriculum Selection (RCS), which constructs a stable, noise-resistant learning schedule through

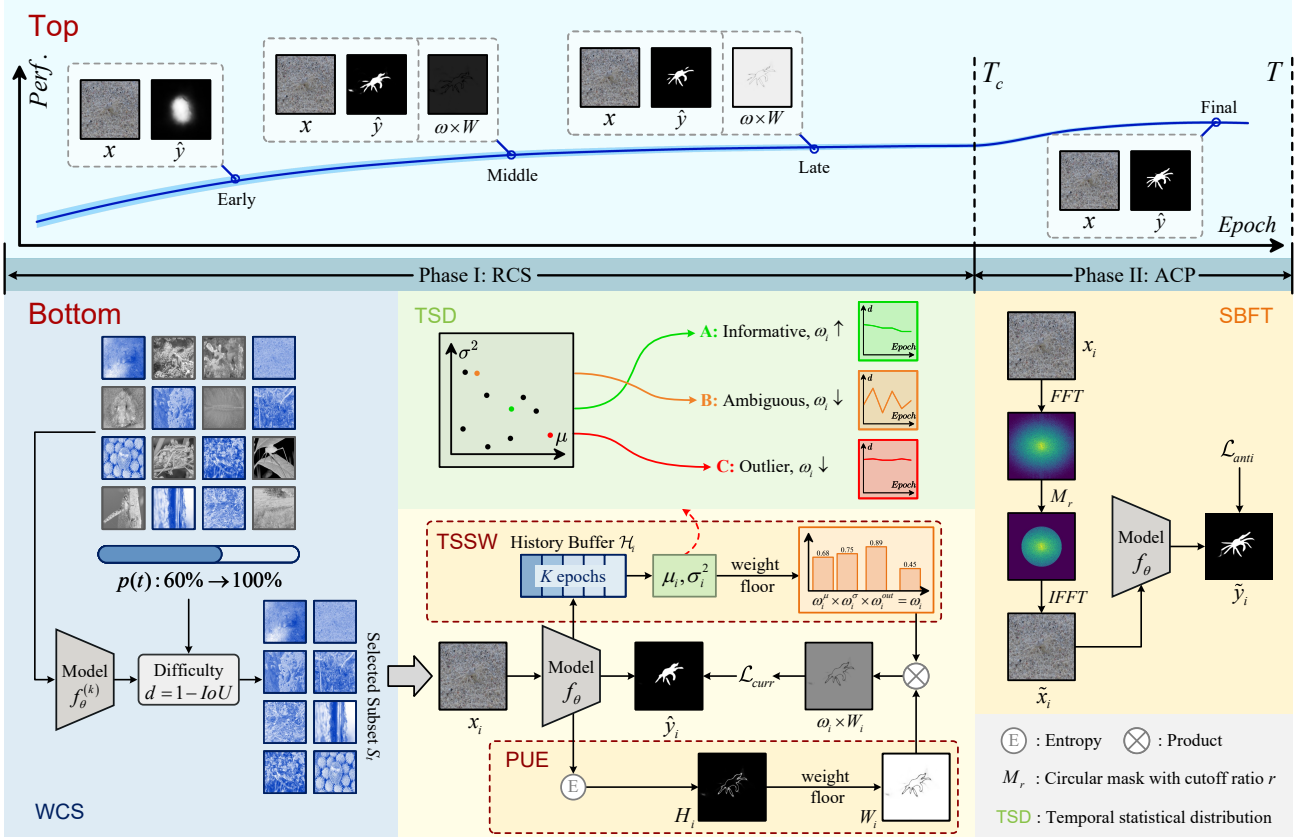


Figure 2. Framework of CurriSeg. In (1), performance improves steadily; brighter regions in  $\omega \times W$  indicate more trustworthy pixels. In (2), Phase I applies Robust Curriculum Selection with sample/pixel-level weighting, while Phase II includes Anti-Curriculum Promotion via spectral filtering. TSD describes how to distinguish hard-but-informative samples from ambiguous or outlier ones.

temporal statistics monitoring and uncertainty-aware weighting, and (2) Anti-Curriculum Promotion (ACP), which deliberately attenuates high-frequency cues to encourage the model to exploit complementary, subtle discriminative cues.

### 3.1. Robust Curriculum Selection

This phase targets the noise–ambiguity dilemma inherent in CECS by monitoring temporal behavior of samples and incorporating pixel-level uncertainty, yielding a joint image- and pixel-level curriculum for stable optimization.

**Warm-up curriculum strategy.** Following (Bengio et al., 2009), we design a warm-up curriculum strategy (WCS) that progressively expands the training set using a model-centric difficulty measure based on segmentation performance. Concretely, we maintain a historical checkpoint  $f_{\theta^{(k)}}$  saved every  $K$  epochs (we set  $K = 10$ , so  $k \in \{0, 10, 20, \dots\}$ ). During the subsequent  $K$  epochs, this checkpoint serves as an evaluator of sample difficulty. For each training sample  $i$ , we compute its difficulty score  $d_i$ , formulated as:

$$d_i = 1 - IoU(f_{\theta^{(k)}}(x_i), y_i), \quad (1)$$

where  $f_{\theta^{(k)}}(x_i)$  is the predicted mask of the concealed input  $x_i$  from checkpoint model,  $y_i$  is the ground-truth, and

$IoU(\cdot)$  is the Intersection-over-Union. Samples with higher  $d_i$  are considered more difficult for the current model.

We then implement a warm-up curriculum schedule by gradually expanding the training set. Let  $p(t)$  denote the selection percentile at epoch  $t$ , and let  $\text{Per}(\{d_n\}_{n=1}^N, p(t))$  be the corresponding percentile threshold over all  $N$  training samples. The active training subset at epoch  $t$  is

$$S_t = \{i \mid d_i \leq \text{Per}(\{d_n\}_{n=1}^N, p(t))\}, \quad (2)$$

$$p(t) = p_{min} + (1 - p_{min}) \cdot \frac{t - 1}{T_c - 1},$$

where  $1 \leq t \leq T_c$  and  $p_{min} = 0.6$ . In practice, training starts from a subset of relatively easy samples (small  $p(t)$ ) and progressively incorporates more difficult ones as  $p(t)$  increases. This encourages the model to first consolidate its understanding of easier cases before being exposed to the full complexity and ambiguity dataset.

**Temporal statistics-based sample weighting.** Within the selected curriculum subset, not all samples contribute equally to effective learning. Due to object concealment, some samples are intrinsically ambiguous or affected by annotation noise; treating them on par with well-defined samples may destabilize optimization. To mitigate this,

**Algorithm 1** CurriSeg: Dual-Phase Curriculum Learning

**Require:** Dataset  $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ , model  $f_\theta$ , epochs  $T$ , curriculum phase  $T_c$ , number of training samples  $N$

**Ensure:** Trained model  $f_\theta$

- 1: Init  $\theta$ , history buffer  $\mathcal{H}$ , checkpoint  $\theta^{(0)} \leftarrow \theta$
- 2: **for**  $t = 1$  to  $T$  **do**
- 3:   **if**  $t \leq T_c$  **then**
- 4:   // Phase 1: Robust Curriculum Selection
- 5:     Compute difficulty  $d_i = 1 - \text{IoU}(f_{\theta^{(k)}}(x_i), y_i)$
- 6:     Select  $\mathcal{S}_t \leftarrow \{i \mid d_i \leq \text{Per}(\{d_n\}_{n=1}^N, p(t))\}$
- 7:     Compute sample weights  $\{\omega_i\}$  from  $(\mu_i, \sigma_i^2)$
- 8:     **for** each batch  $\mathcal{B} \subset \mathcal{S}_t$  **do**
- 9:       Forward:  $\hat{y}_i \leftarrow f_\theta(x_i)$
- 10:       Compute pixel weights  $W_i$  via entropy
- 11:        $\mathcal{L}_{\text{curr}} \leftarrow \frac{1}{|\mathcal{B}|} \sum_i \omega_i \cdot (L_{BCE}^w(W_i) + L_{IoU}^w(W_i))$
- 12:       Update  $\theta, \mathcal{H}$
- 13:     **end for**
- 14:     Save checkpoint if  $t \bmod K = 0$
- 15:   **else**
- 16:   // Phase 2: Anti-Curriculum Promotion
- 17:    **for** each batch  $\mathcal{B} \subset \mathcal{D}$  **do**
- 18:      $\tilde{x}_i \leftarrow \mathcal{F}^{-1}(\mathcal{F}(x_i) \odot M_r)$
- 19:      $\tilde{y}_i \leftarrow f_\theta(\tilde{x}_i)$
- 20:      $\mathcal{L}_{\text{anti}} \leftarrow \frac{1}{|\mathcal{B}|} \sum_i [L_{BCE}^w(\tilde{y}_i, y_i) + L_{IoU}^w(\tilde{y}_i, y_i)]$
- 21:     Update  $\theta$
- 22:    **end for**
- 23:   **end if**
- 24: **end for**
- 25: **return**  $f_\theta$

we exploit temporal statistics of sample-wise performance, termed TSSW, to down-weight problematic samples.

For each sample  $i$ , we maintain a circular buffer storing its difficulty scores over the past  $K$  epochs:  $\mathcal{H}_i = \{d_i^{(1)}, \dots, d_i^{(K)}\}$ , computing two temporal statistics:

$$\mu_i = \frac{1}{K} \sum_{k=1}^K d_i^{(k)}, \quad \sigma_i^2 = \frac{1}{K} \sum_{k=1}^K (d_i^{(k)} - \mu_i)^2. \quad (3)$$

These statistics enable CurriSeg to distinguish hard-but-informative samples from noisy or ambiguous ones:

- **High variance** indicates that the model’s predictions fluctuate significantly over time, suggesting the sample lies near the decision boundary or contains inherent ambiguity that destabilizes learning.
- **High mean error with low variance** indicates consistent failure, signaling a potential outlier or annotation error that the model cannot resolve.

By first applying min-max normalization to obtain  $\tilde{\mu}_i$  and  $\tilde{\sigma}_i^2$ , we convert the temporal statistics into weights:

$$\omega_i^\mu = 1 - \tilde{\mu}_i, \quad \omega_i^\sigma = \exp\left(-\frac{(\tilde{\sigma}_i^2 - \sigma^*)^2}{2\gamma^2}\right), \quad (4)$$

where  $\sigma^* = 0.5$  is the optimal variance level.  $\gamma = 0.2$

controls tolerance to variance deviation. The final weight is:

$$\omega_i = W_{min}^s + (1 - W_{min}^s) \cdot \omega_i^\mu \cdot \omega_i^\sigma \cdot (1 - \tilde{\mu}_i \cdot (1 - \tilde{\sigma}_i^2)), \quad (5)$$

where  $W_{min}^s = 0.1$  provides a weight lower bound and  $\cdot$  is product. We denote the last term as  $\omega_i^{out}$ , i.e.,  $\omega_i^{out} = 1 - \tilde{\mu}_i \cdot (1 - \tilde{\sigma}_i^2)$ .  $\omega_i^{out}$  penalizes the ‘‘outlier’’ pattern of high mean error combined with low variance. Hence, CurriSeg emphasizes hard-but-informative samples while suppressing those that are unstable or likely mislabeled.

**Pixel-level uncertainty estimation.** Segmentation exhibits within-sample variation: boundary and low-contrast pixels are harder than homogeneous ones. If treated equally, gradients from uncertain pixels can dominate optimization. We introduce entropy-based pixel weighting (PUE) to attenuate uncertain pixels while preserving informative gradients.

In contrast to prior work that typically uses uncertainty to filter pseudo labels (He et al., 2024a; 2025a;d), we apply uncertainty directly on the predicted mask. The goal is not to hide failure regions from the model, but to prevent them from disproportionately influencing learning.

For a prediction logit  $\hat{y}_{h,w}$  at spatial location  $(h, w)$ , we compute its predicted probability  $p_{h,w}$  with Sigmoid  $\sigma$ :  $p_{h,w} = \sigma(\hat{y}_{h,w})$ . The normalized prediction entropy is:

$$H_{h,w} = -p_{h,w} \log_2 p_{h,w} - (1 - p_{h,w}) \log_2 (1 - p_{h,w}). \quad (6)$$

$H_{h,w} \in [0, 1]$  reaches its maximum when  $p_{h,w} = 0.5$  (maximum uncertainty) and minimum when  $p_{h,w} \in \{0, 1\}$  (complete certainty). Instead of hard masking, we define a soft weighting matrix  $W_i$  whose entry at  $(h, w)$  and epoch  $t$  is:

$$W_{h,w}(t) = W_{min} + (1 - W_{min}) \cdot (1 - \beta(t) \cdot H_{h,w}), \quad (7)$$

where  $W_{min} = 0.1$  sets a non-zero weight floor, and  $\beta(t) = (1 - t/T_c)$  is a curriculum-aware coefficient that decays over the curriculum phase of length  $T_c$ . Early in training, high-entropy pixels receive substantially reduced weights, limiting gradient noise from ambiguous regions. As training progresses and  $\beta(t)$  decreases, full supervision is gradually restored, while the weight floor ensures continuous, albeit reduced, gradient flow from all pixels.

**Unified curriculum loss.** Our overall objective follows standard CECS practice (Fan et al., 2020; He et al., 2023a), augmented with the proposed curriculum mechanisms. For a sample  $i \in \mathcal{S}_t$  at epoch  $t$ , the loss is defined as:

$$\mathcal{L}_{\text{curr}} = \omega_i \cdot [L_{BCE}^w(\hat{y}_i, y_i, W_i) + L_{IoU}^w(\hat{y}_i, y_i, W_i)], \quad (8)$$

where  $\omega_i$  is the sample-level weight from temporal statistics.  $L_{BCE}^w$  and  $L_{IoU}^w$  are the weighted binary cross-entropy loss and weighted intersection-over-union loss with the combined pixel-level weight  $W_i$ . This joint image- and pixel-level curriculum yields a stable learning schedule.

### 3.2. Anti-Curriculum Promotion

Once the model has reached a stable representation regime, we introduce an Anti-Curriculum Promotion phase to en-

Table 1. Results on camouflaged object detection. The suffix “+” indicates that the network is trained under our CurriSeg.  $\Delta$ : increase.

Methods	Backbones	CHAMELEON				CAMO				COD10K				NC4K				COD $\Delta$ (%)
		$M \downarrow$	$F_{\beta} \uparrow$	$E_{\phi} \uparrow$	$S_{\alpha} \uparrow$	$M \downarrow$	$F_{\beta} \uparrow$	$E_{\phi} \uparrow$	$S_{\alpha} \uparrow$	$M \downarrow$	$F_{\beta} \uparrow$	$E_{\phi} \uparrow$	$S_{\alpha} \uparrow$	$M \downarrow$	$F_{\beta} \uparrow$	$E_{\phi} \uparrow$	$S_{\alpha} \uparrow$	
SINet (Fan et al., 2020)	ResNet50	0.034	0.823	0.936	0.872	0.092	0.712	0.804	0.745	0.043	0.667	0.864	0.776	0.058	0.768	0.871	0.808	—
MGL-R (Zhai et al., 2021)	ResNet50	0.031	0.825	0.917	0.891	0.088	0.738	0.812	0.775	0.035	0.680	0.851	0.814	0.053	0.778	0.867	0.833	—
PreyNet (Zhang et al., 2022a)	ResNet50	0.027	0.844	0.948	0.895	0.077	0.763	0.854	0.790	0.034	0.715	0.894	0.813	0.047	0.798	0.887	0.838	—
FEDER (He et al., 2023a)	ResNet50	0.028	0.850	0.944	0.892	0.070	0.775	0.870	0.802	0.032	0.715	0.892	0.810	0.046	0.808	0.900	0.842	—
FEDER+ (Ours)	ResNet50	0.026	0.858	0.952	0.898	0.068	0.790	0.881	0.807	0.030	0.736	0.910	0.818	0.043	0.825	0.912	0.850	2.46 $\uparrow$
FSEL (Sun et al., 2024)	ResNet50	0.029	0.847	0.941	0.893	0.069	0.779	0.881	0.816	0.032	0.722	0.891	0.822	0.045	0.807	0.901	0.847	—
FSEL+ (Ours)	ResNet50	0.028	0.856	0.949	0.898	<b>0.067</b>	<b>0.792</b>	<b>0.889</b>	<b>0.819</b>	0.030	0.742	0.909	0.831	0.042	0.823	0.919	0.855	2.22 $\uparrow$
RUN (He et al., 2025e)	ResNet50	0.027	0.855	0.952	0.895	0.070	0.781	0.868	0.806	0.030	0.747	0.903	0.827	0.042	0.824	0.908	0.851	—
RUN+ (Ours)	ResNet50	<b>0.025</b>	<b>0.863</b>	<b>0.960</b>	<b>0.900</b>	0.068	<b>0.801</b>	0.879	0.813	<b>0.029</b>	<b>0.765</b>	<b>0.917</b>	<b>0.836</b>	<b>0.040</b>	<b>0.835</b>	<b>0.921</b>	<b>0.858</b>	2.13 $\uparrow$
BSA-Net (Zhu et al., 2022)	Res2Net50	0.027	0.851	0.946	0.895	0.079	0.768	0.851	0.796	0.034	0.723	0.891	0.818	0.048	0.805	0.897	0.841	—
RUN (He et al., 2025e)	Res2Net50	0.024	0.879	0.956	0.907	0.066	0.815	0.905	0.843	0.028	0.764	0.914	0.849	0.041	0.830	0.917	0.859	—
RUN+ (Ours)	Res2Net50	<b>0.023</b>	<b>0.891</b>	<b>0.963</b>	<b>0.911</b>	<b>0.065</b>	<b>0.820</b>	<b>0.912</b>	<b>0.845</b>	<b>0.026</b>	<b>0.785</b>	<b>0.933</b>	<b>0.857</b>	<b>0.038</b>	<b>0.852</b>	<b>0.932</b>	<b>0.870</b>	2.23 $\uparrow$
CamoDiff (Sun et al., 2025)	PVT V2	0.022	0.868	0.952	0.908	0.042	0.853	0.936	0.878	0.019	0.815	0.943	0.883	0.028	0.858	0.942	0.895	—
RUN (He et al., 2025e)	PVT V2	0.021	0.877	0.958	0.916	0.045	0.861	0.934	0.877	0.021	0.810	0.941	0.878	0.030	0.868	0.940	0.892	—
RUN+ (Ours)	PVT V2	<b>0.019</b>	<b>0.893</b>	<b>0.971</b>	<b>0.922</b>	<b>0.042</b>	<b>0.879</b>	<b>0.952</b>	<b>0.885</b>	<b>0.018</b>	<b>0.828</b>	<b>0.957</b>	<b>0.886</b>	<b>0.026</b>	<b>0.889</b>	<b>0.958</b>	<b>0.903</b>	3.94 $\uparrow$

Table 2. Training overhead analysis (batch size: 2).

Metrics	FEDER	FEDER+	FSEL	FSEL+	RUN	RUN+
Training Time (h)	9.62	6.84	11.54	5.96	12.64	8.32
GPU Mem (G)	1.53	1.62	2.83	2.92	3.66	3.75
Perf. Gain (%)	—	2.46 $\uparrow$	—	2.22 $\uparrow$	—	2.13 $\uparrow$

Table 3. Results on polyp image segmentation.

Methods	CVC-ColonDB			ETIS			PIS $\Delta$ (%)
	mDice $\uparrow$	mIoU $\uparrow$	$S_{\alpha}$ $\uparrow$	mDice $\uparrow$	mIoU $\uparrow$	$S_{\alpha}$ $\uparrow$	
PolypPVT (Dong et al., 2023)	0.808	0.727	0.865	0.787	0.706	0.871	—
CoInNet (Jain et al., 2023)	0.797	0.729	0.875	0.759	0.690	0.859	—
LSSNet (Wang et al., 2024)	0.820	0.741	0.867	0.779	0.701	0.867	—
MEGANet (Bui et al., 2024)	0.793	0.714	0.854	0.739	0.665	0.836	—
MEGANet+ (Ours)	0.815	0.736	0.866	0.755	0.682	0.848	2.24 $\uparrow$
RUN (He et al., 2025e)	0.822	0.742	0.880	0.788	0.709	0.878	—
RUN+ (Ours)	<b>0.845</b>	<b>0.763</b>	<b>0.893</b>	<b>0.806</b>	<b>0.720</b>	<b>0.890</b>	2.05 $\uparrow$

hance its performance on hard samples. This phase counteracts the tendency of the network to exploit superficial texture shortcuts, an inherent bias in dense prediction, and encourages the discovery of complementary, subtler cues.

We propose Spectral-Blindness Fine-Tuning (SBFT), which attenuates high-frequency input components. By suppressing texture information, SBFT forces reliance on low-frequency structural patterns and contextual semantics, promoting extraction of intrinsic, task-relevant signals.

Given the concealed input  $x_i$ , we apply a low-pass filter in the frequency domain, formulated as:

$$\tilde{x}_i = \mathcal{F}^{-1}(\mathcal{F}(x_i) \odot M_r), \quad (9)$$

where  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  denote the 2D Fourier transform and its inverse, and  $M_r$  is a circular mask that preserves frequency components within a radius ratio  $r$ , defined as:

$$M_r(u, v) = \mathbf{1} \left[ (u - c_u)^2 + (v - c_v)^2 \leq \frac{r \cdot \min(H, W)}{2} \right], \quad (10)$$

where  $\mathbf{1}(\cdot)$  is a indicator function,  $H, W$  are the image height and width,  $(c_u, c_v)$  is the frequency domain center and  $r = 0.95$  controls the cutoff ratio. By attenuating high-frequency texture details, SBFT creates an information bottleneck that reduces reliance on superficial patterns; the model is instead driven to exploit low-frequency boundaries, semantic context, and subtle intensity variations, which tend to be more robust across diverse concealed scenarios and

Table 4. Results on transparent object detection.

Methods	GDD			GSD			TOD $\Delta$ (%)
	mIoU $\uparrow$	$F_{\beta}^{max}$ $\uparrow$	$M \downarrow$	mIoU $\uparrow$	$F_{\beta}^{max}$ $\uparrow$	$M \downarrow$	
EBLNet (He et al., 2021)	0.870	0.922	0.064	0.817	0.878	0.059	—
IEBAF (Han et al., 2024)	0.887	0.944	0.056	0.861	0.926	0.049	—
GhostingNet (Yan et al., 2024)	0.893	0.943	0.054	0.838	0.904	0.055	—
RFENet (Fan et al., 2023b)	0.886	0.938	0.057	0.865	0.931	0.048	—
RFENet+ (Ours)	0.893	0.952	0.052	0.875	0.939	0.045	3.22 $\uparrow$
RUN (He et al., 2025e)	0.895	0.952	0.051	0.866	0.938	0.043	—
RUN+ (Ours)	<b>0.902</b>	<b>0.966</b>	<b>0.047</b>	<b>0.878</b>	<b>0.945</b>	<b>0.039</b>	3.59 $\uparrow$

thus enhance the ability of the model in handling hard cases.

Let  $\tilde{y}$  denote the predicted mask for  $\tilde{x}_i$ . The loss function is:

$$\mathcal{L}_{anti} = L_{BCE}^w(\tilde{y}_i, y_i) + L_{IoU}^w(\tilde{y}_i, y_i). \quad (11)$$

### 3.3. Overall Training Pipeline

Algorithm 1 summarizes the CurriSeg training procedure:

- **Phase 1 (Epochs 1 to  $T_c$ ):** Robust Curriculum Selection builds stable representations through progressive sample introduction, temporal statistics-based weighting, and pixel-level uncertainty estimation.
- **Phase 2 (Epochs  $T_c$  to  $T$ ):** Anti-Curriculum Promotion enhances robustness to hard cases through SBFT.

CurriSeg does not modify the model architecture and introduces no extra parameters at inference. The overall training cost remains comparable to standard schedules, while the curriculum phase often accelerates effective convergence by focusing computation on more informative samples.

## 4. Experiments

**Experimental setup.** Our CurriSeg framework is implemented in PyTorch and trained on two NVIDIA RTX 4090 GPUs, using the Adam optimizer with momentum terms parameters  $(\beta_1, \beta_2) = (0.9, 0.999)$ . We adopt an initial warm-up period of 10 epochs in which all training samples are used without curriculum filtering, allowing the model to acquire a basic representation capacity before difficulty-based selection is applied. Unless specified, the curriculum phase length is set to  $T_c = 60$  epochs, followed by an anti-curriculum fine-tuning phase up to  $T = 70$  epochs. When

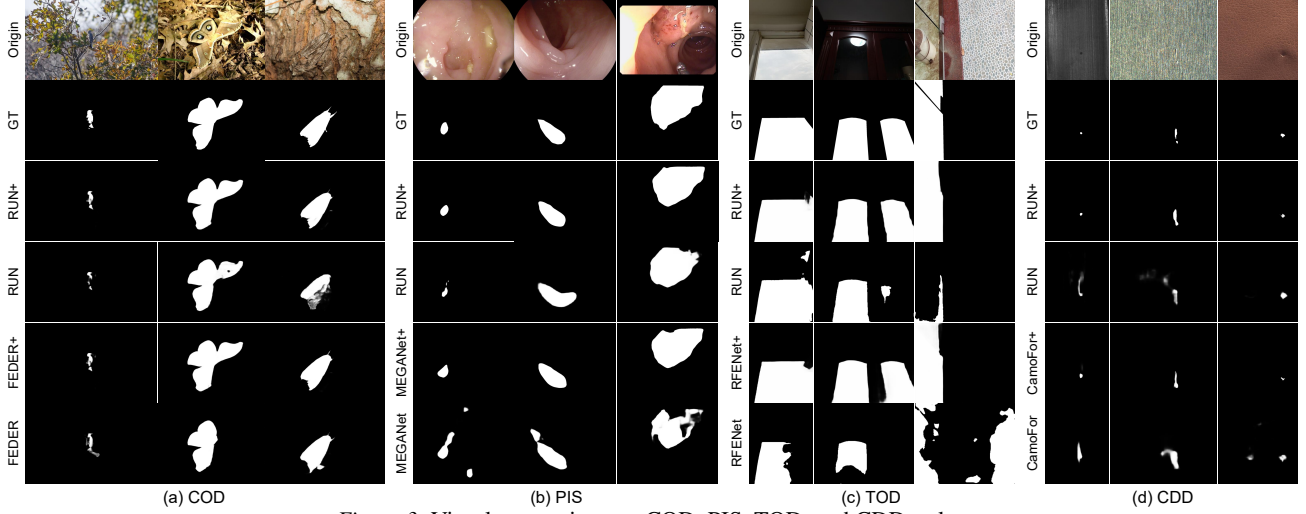


Figure 3. Visual comparison on COD, PIS, TOD, and CDD tasks.

Table 5. Results on concealed defect detection.

Methods	$S_\alpha \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$F_\beta \uparrow$	$F_\beta^{mean} \uparrow$	$\Delta (\%)$
SiNet V2 (Fan et al., 2021a)	0.551	0.102	0.567	0.223	0.248	—
HitNet (Hu et al., 2023)	0.563	0.118	0.564	0.298	0.298	—
OAFoFormer (Yang et al., 2023)	0.541	0.121	0.535	0.216	0.239	—
CamoFormer (Yin et al., 2024)	0.589	0.100	0.588	0.330	0.329	—
CamoFormer+ (Ours)	<b>0.599</b>	0.091	0.603	<b>0.345</b>	<b>0.346</b>	4.59 $\uparrow$
RUN (He et al., 2025e)	0.590	0.068	0.595	0.298	0.299	—
RUN+ (Ours)	0.598	<b>0.061</b>	<b>0.612</b>	0.309	0.310	4.38 $\uparrow$

Table 6. Breakdown ablations of CurriSeg.

WCS	PUE	TSSW	SBFT	$M \downarrow$	$F_\beta \uparrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$
×	×	×	×	0.032	0.715	0.892	0.810
✓	×	×	×	0.035	0.697	0.870	0.801
✓	✓	×	×	0.033	0.718	0.895	0.809
✓	✓	✓	×	0.031	0.729	0.904	0.815
×	×	×	✓	0.031	0.723	0.902	0.813
✓	✓	✓	✓	<b>0.030</b>	<b>0.736</b>	<b>0.910</b>	<b>0.818</b>

Table 7. Ablation study of the robust curriculum selection component in CurriSeg. We evaluate performance in the COD task on *COD10K*. SPL and TSC are shorts for self-paced learning and teacher-student curriculum. “w/o” indicates without.

Metrics	WCS				TSSW					PUE		CurriSeg (Ours)
	SPL $\rightarrow$ WCS	TSC $\rightarrow$ WCS	$p_1(t) \rightarrow p(t)$	$p_2(t) \rightarrow p(t)$	w/o $\omega_i^H$	w/o $\omega_i^\sigma$	w/o $\omega_i^{\sigma_{out}}$	$\omega_i^{\sigma_1} \rightarrow \omega_i^\sigma$	$\omega_i^{\sigma_2} \rightarrow \omega_i^\sigma$	w/o $\beta(t)$	$\beta_1(t) \rightarrow \beta(t)$	
$M \downarrow$	<b>0.030</b>	<b>0.030</b>	0.031	0.031	0.031	0.031	0.030	0.031	0.031	0.031	0.030	<b>0.030</b>
$F_\beta \uparrow$	<b>0.738</b>	0.735	0.730	0.726	0.729	0.727	0.732	0.730	0.727	0.730	0.732	0.736
$E_\phi \uparrow$	0.908	<b>0.910</b>	0.905	0.902	0.903	0.904	0.906	0.900	0.903	0.900	0.907	<b>0.910</b>
$S_\alpha \uparrow$	<b>0.820</b>	0.817	0.815	0.812	0.814	0.810	0.810	0.808	0.809	0.814	0.815	0.818

Table 8. Ablation study of anti-curriculum promotion in CurriSeg. TAL and AA are texture-aware loss and aggressive augmentation.

Metrics	Hard-only SBFT	Random SBFT	Square filter	Progressive filter	Blur $\rightarrow$ SBFT	Noise $\rightarrow$ SBFT	TAL $\rightarrow$ SBFT	AA $\rightarrow$ SBFT	Reverse CurriSeg	CurriSeg
$M \downarrow$	0.031	0.031	0.031	<b>0.030</b>	0.033	0.032	0.032	0.033	0.052	<b>0.030</b>
$F_\beta \uparrow$	0.730	0.728	0.729	0.734	0.710	0.719	0.725	0.718	0.632	<b>0.736</b>
$E_\phi \uparrow$	0.899	0.902	0.905	<b>0.910</b>	0.875	0.866	0.897	0.868	0.787	<b>0.910</b>
$S_\alpha \uparrow$	0.813	0.812	0.814	0.816	0.804	0.806	0.808	0.807	0.742	<b>0.818</b>

integrating CurriSeg with existing CECS backbones, all input images are resized to  $352 \times 352$  during both training and testing, and all other hyperparameters follow the original configurations of the corresponding backbone methods.

#### 4.1. Comparative Evaluation

We conduct extensive experiments across CECS tasks, with datasets and metrics detailed in Sec. A.1 in the appendix. All results follow standardized evaluation protocols.

**Camouflaged object detection.** As shown in Table 1, integrating our CurriSeg into cutting-edge methods (denoted by “+”) consistently improves performance across ResNet50 (He et al., 2016), Res2Net50 (Gao et al., 2019), and PVT V2 (Wang et al., 2022). Moreover, the qualitative results in Fig. 3 demonstrate that models optimized under our CurriSeg yield more accurate segmentation, even in

highly challenging scenarios. Together, these results substantiate the effectiveness of our dual-phase curriculum strategy. Beyond performance gains, CurriSeg also improves training efficiency. As reported in Table 2, training time is reduced by 28.9% -48.4% across backbones, since curriculum selection concentrates computation on informative samples rather than all data. GPU memory overhead is negligible ( $\leq 0.1G$ ), since no extra parameters are introduced. Hence, our CurriSeg is more efficient than the standard manner.

**Polyp image segmentation.** Experiments are conducted on the PIS task using *CVC-ColonDB* and *ETIS*. In line with RUN (He et al., 2025e), PVT V2 is employed as the backbone. As evidenced by the quantitative results in Table 3 and the visualizations in Fig. 3, models trained under our CurriSeg consistently outperform their original versions.

**Transparent object detection.** As reported in Table 4 and Fig. 3, CurriSeg integration yields consistent gains over

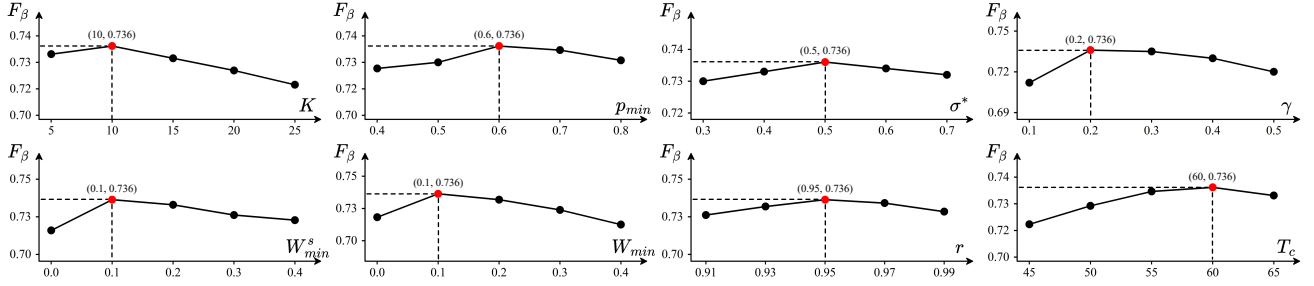


Figure 4. Sensitivity analysis of hyperparameters, including  $K$ ,  $p_{min}$ ,  $\sigma^*$ ,  $\gamma$ ,  $W_{min}^s$ ,  $W_{min}$ ,  $r$ , and  $T_c$ .

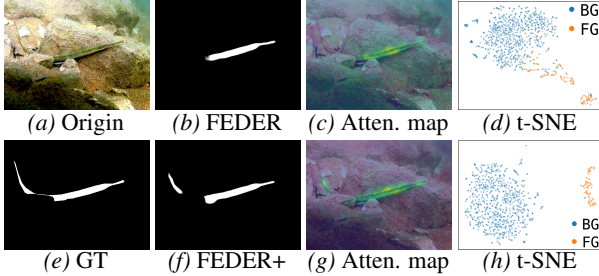


Figure 5. Visualization of attention map and t-SNE of foreground (orange) and background (blue) features. (b)-(d) and (f)-(h) correspond to the prediction, attention map, and t-sne of FEDER trained under the standard setting and our CurriSeg framework.

Table 9. Generalization of CurriSeg on other settings, including weak supervision with scribble on *COD10K* (WS-SAM (He et al., 2024a) and SEE (He et al., 2025a)), semi-supervision with 1/16 labeled data on *COD10K* (CoSOD (Chakraborty et al., 2024) and SEE), multi-modality with depth on *COD10K* (DSAM (Yu et al., 2024) and MultiCOS (Fang et al., 2025)), and video segmentation on *CAD* (Cheng et al., 2022b) (STL-Net (Cheng et al., 2022b) and ZoomNext (Pang et al., 2024)). Metrics are reported according to each setting.

(a) Weak supervision.					(b) Semi-supervision.					(c) Multi-modality.					(d) Video segmentation.				
Methods	$M \downarrow$	$F_\beta \uparrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	Methods	$M \downarrow$	$F_\beta \uparrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	Methods	$M \downarrow$	$F_\beta^{max} \uparrow$	$E_\phi^{max} \uparrow$	$S_\alpha \uparrow$	Methods	$M \downarrow$	$F_\beta \uparrow$	$E_\phi^{max} \uparrow$	$S_\alpha \uparrow$
WS-SAM	0.038	0.719	0.878	0.803	CoSOD	0.055	0.650	0.795	0.740	DSAM	0.033	0.807	0.931	0.846	STL-Net	0.030	0.481	0.845	0.696
WS-SAM+	0.035	0.733	0.886	0.813	CoSOD+	0.051	0.672	0.810	0.749	DSAM+	0.030	0.821	0.942	0.853	STL-Net+	0.026	0.505	0.861	0.706
SEE	0.036	0.729	0.883	0.807	SEE	0.046	0.679	0.803	0.745	MultiCOS	0.020	0.950	0.946	0.880	ZoomNext	0.020	0.593	0.865	0.757
SEE+	<b>0.033</b>	<b>0.738</b>	<b>0.893</b>	<b>0.816</b>	SEE+	<b>0.043</b>	<b>0.695</b>	<b>0.817</b>	<b>0.757</b>	MultiCOS+	<b>0.018</b>	<b>0.958</b>	<b>0.953</b>	<b>0.883</b>	ZoomNext+	<b>0.018</b>	<b>0.611</b>	<b>0.873</b>	<b>0.761</b>

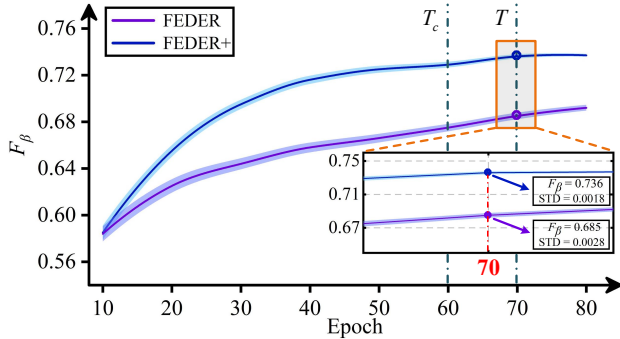


Figure 7. Performance ( $F_\beta$ ) vs. training epoch on *COD10K*. Phase 1:  $0-T_c$  and Phase 2:  $T_c-T$ . The initial 10 epochs are used for warm-up, where FEDER shares the same training details with FEDER+, and thus be omitted for clarification.

existing methods, highlighting its utility as a general performance enhancer for perception in autonomous driving.

**Concealed defect detection.** We validate the generalizability of models trained with CurriSeg on CDD. Models pre-trained on COD are directly transferred to the *CDS2K* dataset for defect segmentation, following RUN (He et al., 2025e). As shown in Table 5 and Fig. 3, models optimized under CurriSeg consistently achieve superior performance,

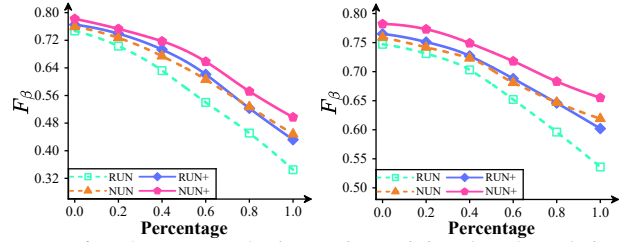


Figure 6. Robustness under increasing training data degradation. **Left:** High-frequency information loss (blur, low-light, haze). **Right:** High-frequency interference (noise). CurriSeg maintains superior performance across all degradation ratios, indicating the potential of our framework in real-world degradation scenarios.

demonstrating enhanced cross-domain generalization.

## 4.2. Ablation Study

We conduct ablations on *COD10K* using FEDER, a cutting-edge network that integrates frequency guidance and attention mechanisms, as the baseline in Secs. 4.2 and 4.3.

**Effect of robust curriculum selection.** As shown in Table 6, robust curriculum selection (WCS+PUE+TSSW) stabilizes training and suppresses ambiguous supervision. In Table 7, replacing WCS with self-paced learning (Kumar et al., 2010) or teacher-student curricula (Matiisen et al., 2019) yields comparable results but adds complexity. Altering the warm-up schedule from  $P(t)$  to  $p_1(t) = p_{min} + (1 - p_{min}) \cdot (\frac{t-1}{T_c-1})^2$  or  $p_2(t) = p_{min} + (1 - p_{min}) \cdot (\frac{t-1}{T_c-1})^{0.5}$  degrades performance, suggesting  $p(t)$  offers a better stability-coverage trade-off. Within TSSW, dropping any component ( $\omega_i^\mu$ ,  $\omega_i^\sigma$ , and  $\omega_i^{out}$ ) or replacing  $\omega_i^\sigma$  with  $\omega_i^{\sigma^1} = \max(0, 1 - \frac{|\bar{\sigma}_i^2 - \sigma^*|}{\gamma})$  and  $\omega_i^{\sigma^2} = 1 - (\frac{\bar{\sigma}_i^2 - \sigma^*}{\gamma})^2$  leads to noticeable drops. Finally, removing the entropy-based pixel reweighting  $\beta(t)$  or using exponential decay  $\beta_1(t) = \exp(-\frac{t}{T_c})$  also hurts.

**Effect of anti-curriculum promotion.** As shown in Ta-

Table 10. Generalization of CurriSeg on other dense prediction tasks with cutting-edge methods, including semantic segmentation on ADE20K (Zhou et al., 2017) (PEM (Cavagnero et al., 2024) and CGRSeg (Ni et al., 2024)), instance segmentation on COCO (Lin et al., 2014) (M2Form and FasIns are short for Mask2Former (Cheng et al., 2022a) and FastInst (He et al., 2023b)), infrared small target detection on IRSTD-1k (Zhang et al., 2022b) (ISNet (Zhang et al., 2022b) and IRSAM (Zhang et al., 2024)), and shadow detection on SBU (Vicente et al., 2016) (SARA (Sun et al., 2023) and spider (Zhao et al., 2024)), with BER (Vicente et al., 2015) used for evaluation.

(a) Semantic segmentation.				(b) Instance segmentation.				(c) Infrared small target detection.				(d) Shadow detection.							
Metric	PEM	PEM+	CGRSeg	CGRSeg+	Metric	M2Form	M2Form+	FasIns	FasIns+	Metric	ISNet	ISNet+	IRSAM	IRSAM+	Metric	SARA	SARA+	Spider	Spider+
IoU $\uparrow$	45.0	45.9	45.5	<b>46.8</b>	AP $\uparrow$	38.0	38.6	38.6	<b>39.1</b>	IoU $\uparrow$	68.77	70.62	73.69	<b>75.52</b>	BER $\downarrow$	0.043	0.040	0.040	<b>0.038</b>

Table 11. Compatibility of our CurriSeg with advanced architectures (attention-based FSEL (Sun et al., 2024), multiscale-based ZoomNet (Pang et al., 2022), and uncertainty-based UGTR (Yang et al., 2021)) and foundation models (SAM-adaptor (SAM-a) (Chen et al., 2023), SAM2-adaptor (SAM2-a) (Chen et al., 2024), and SAM3-adaptor (SAM3-a) (Chen et al., 2025)) on COD10K. In (b)-(d), the suffix “\*” means modifying the general CurriSeg framework with simple, architecture-specific changes. The further performance improvement indicates the potential of our CurriSeg in facilitating the community.

(a)	(b) Attention.			(c) Multiscale.			(d) Uncertainty.			(e) Foundation model adapters.					
Metr.	FSEL	FSEL+	FSEL*	ZoomNet	ZoomNet+	ZoomNet*	UGTR	UGTR+	UGTR*	SAM-a	SAM-a+	SAM2-a	SAM2-a+	SAM3-a	SAM3-a+
$M \downarrow$	0.032	0.030	<b>0.029</b>	0.029	0.028	<b>0.027</b>	0.036	0.034	<b>0.032</b>	0.025	0.023	0.018	0.016	0.015	<b>0.014</b>
$F_\beta \downarrow$	0.722	0.742	<b>0.749</b>	0.740	0.753	<b>0.758</b>	0.670	0.688	<b>0.706</b>	0.800	0.822	0.848	0.869	0.883	<b>0.892</b>
$E_\phi \uparrow$	0.891	0.909	<b>0.916</b>	0.888	0.898	<b>0.907</b>	0.852	0.865	<b>0.889</b>	0.918	0.935	0.950	0.962	0.965	<b>0.968</b>
$S_\alpha \uparrow$	0.822	0.831	<b>0.834</b>	0.838	0.842	<b>0.845</b>	0.817	0.825	<b>0.830</b>	0.883	0.888	0.899	0.903	0.927	<b>0.930</b>

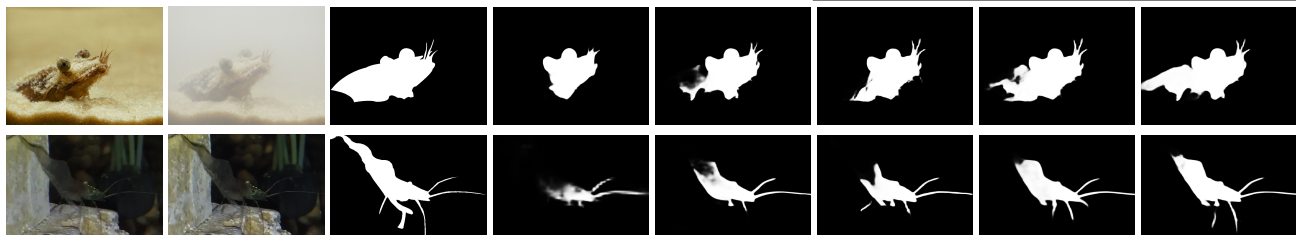


Figure 8. Robustness analysis under high-frequency loss (Row 1: haze) and interference (Row 2: noise). While converged FEDER (f) overfit fragile texture shortcuts, CurriSeg preserves structural integrity by balancing low-frequency context with task-relevant details.

ble 8, applying SBFT only to hard samples (top 50%) or to a random 50% subset falls below CurriSeg. Replacing our circular low-pass mask with square or progressive filters worsens results. with Gaussian blur, additive noise, texture-aware loss, or aggressive augmentation degrades metrics, confirming the benefit of principled spectral design. Reversing the phase order causes dramatic performance drops.

**Hyper-parameter sensitivity.** Fig. 4 shows CurriSeg’s sensitivity to key hyper-parameters. Optimal settings are:  $K = 10$ ,  $p_{min} = 0.6$  (WCS);  $\sigma^* = 0.5$ ,  $\gamma = 0.2$ ,  $W_{min}^s = 0.1$  (TSSW);  $W_{min} = 0.1$  (PUE);  $r = 0.95$ ,  $T_c = 60$  (ACP). The framework is robust within moderate ranges.

### 4.3. Further Analysis and Applications

**Feature space analysis.** As shown in Fig. 5, we visualize attention maps and feature distributions via t-SNE (Maaten & Hinton, 2008). The baseline exhibits substantial cluster overlap, while CurriSeg training yields improved intra-class compactness and inter-class separation, confirming that our CurriSeg induces more discriminative representations.

**Broad generalization across settings and tasks.** We evaluate CurriSeg under weak/semi-supervision, multi-modal learning, and video segmentation (Table 9), as well as semantic/instance segmentation, infrared target detection, and shadow detection (Table 10). Consistent improvements

demonstrate that our paradigm provides a general training principle for dense prediction, highlighting our potential.

**Performance dynamics across epochs.** Fig. 7 shows that CurriSeg consistently surpasses baselines throughout training (averaged over five runs), with faster convergence and reduced variance indicating more stable optimization.

**Robustness under image degradation.** We evaluate robustness by progressively introducing degraded samples (from NUN (He et al., 2025f) with ratios from 0% to 100%). As shown in Fig. 6, CurriSeg outperforms baselines, with the gap widening at higher degradation levels, demonstrating enhanced reliance on low-frequency structural cues.

**Compatibility with advanced architectures.** CurriSeg is architecture-agnostic. Table 11 shows consistent gains on attention-based (FSEL), multi-scale (ZoomNet), uncertainty-aware (UGTR), and foundation model adapters (SAM series). “+”: CurriSeg; “\*”: simple and specific adaptations (Sec. A.2.1), highlighting broad applicability.

## 5. Discussion

CurriSeg offers a frequency-aware perspective extending beyond CECS. (i) **Spectral learning dynamics:** The “stabilize-then-perturb” trajectory first anchors the model on reliable representations through noise-resistant curriculum.

(ii) **Mitigating high-frequency dependency:** SBFT’s effectiveness suggests standard training drifts toward texture shortcuts. As shown in Fig. 8, fully converged baselines often overfit fragile high-frequency cues, leading to collapse under spectral degradations. (iii) **Robustness via information bottlenecks:** By attenuating high-frequency components, CurriSeg enforces reliance on low-frequency boundaries and contextual semantics—a powerful, architecture-agnostic complement to frequency-aware modules.

## 6. Conclusions

We proposed CurriSeg, which follows a “stabilize-then-perturb” process: robust curriculum selection uses temporal loss statistics and pixel-level uncertainty to construct a stable, noise-resistant learning schedule, and an anti-curriculum phase applies SBFT to attenuate high-frequency texture cues. This encourages reliance on structural and contextual information and reduces shortcut behavior. Experiments show consistent performance and robustness gains without extra parameters or training overhead, highlighting our importance for entangled and dense visual scenes.

## Acknowledgment

This work was supported in part by the Foundation Fighting Blindness (BR-CL-0621-0812-DUKE and PPA-1224-0890-DUKE) and Research to Prevent Blindness Unrestricted Grant to Duke University. The first two authors contribute equally to this paper.

## Impact Statement

This work advances segmentation for concealed and context-entangled objects, with potential benefits in medical image analysis and autonomous perception. As with other segmentation systems, deployment in safety-critical settings such as clinical diagnosis or driving requires careful validation, since segmentation errors may carry real-world consequences. We do not foresee additional ethical concerns beyond those common to general dense-prediction research.

## References

- Bengio, Y., Louradour, J., Collobert, R., and Weston, J. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009. 1, 2, 3
- Bui, N.-T., Hoang, D.-H., Nguyen, Q.-T., Tran, M.-T., and Le, N. Meganet: Multi-scale edge-guided attention network for weak boundary polyp segmentation. In *WACV*, pp. 7985–7994, 2024. 5
- Cavagnero, N., Rosi, G., Cuttano, C., Pistilli, F., Ciccone, M., Averta, G., and Cermelli, F. Pem: Prototype-based efficient maskformer for image segmentation. In *CVPR*, pp. 15804–15813, 2024. 8
- Chakraborty, S., Naha, S., Bastan, M., Samaras, D., et al. Unsupervised and semi-supervised co-salient object detection via segmentation frequency statistics. In *WACV*, pp. 332–342, 2024. 7
- Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., and Zhang, S. Sam-adapter: Adapting segment anything in underperformed scenes. In *ICCV*, pp. 3367–3375, 2023. 8
- Chen, T., Lu, A., Zhu, L., Ding, C., Yu, C., Ji, D., Li, Z., Sun, L., Mao, P., and Zang, Y. Sam2-adapter: Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. *arXiv preprint arXiv:2408.04579*, 2024. 8
- Chen, T., Cao, R., Yu, X., Zhu, L., Ding, C., Ji, D., Chen, C., Zhu, Q., Xu, C., Mao, P., et al. Sam3-adapter: Efficient adaptation of segment anything 3 for camouflage object segmentation, shadow detection, and medical image segmentation. *arXiv preprint arXiv:2511.19425*, 2025. 8
- Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., and Girshick, R. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1290–1299, 2022a. 8
- Cheng, X., Xiong, H., Fan, D.-P., Zhong, Y., Harandi, M., Drummond, T., and Ge, Z. Implicit motion handling for video camouflaged object detection. In *CVPR*, pp. 13864–13873, 2022b. 7
- Dong, B., Wang, W., Fan, D.-P., Li, J., Fu, H., and Shao, L. Polyp-pvt: Polyp segmentation with pyramid vision transformers. *CAAI Artif. Intell. Res.*, 2, 2023. 5
- Fan, D.-P., Cheng, M.-M., Liu, Y., and Li, T. Structure-measure: A new way to evaluate foreground maps. In *ICCV*, pp. 4548–4557, 2017. 13
- Fan, D.-P., Ji, G.-P., Sun, G., Cheng, M.-M., and Shen, J. Camouflaged object detection. In *CVPR*, pp. 2777–2787, 2020. 1, 2, 4, 5, 13
- Fan, D.-P., Ji, G.-P., Cheng, M.-M., and Shao, L. Concealed object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021a. 2, 6, 13
- Fan, D.-P., Ji, G.-P., Qin, X., and Cheng, M.-M. Cognitive vision inspired object segmentation metric and loss function. *Scientia Sinica Informationis*, 6(6), 2021b. 13
- Fan, D.-P., Ji, G.-P., Xu, P., Cheng, M.-M., Sakaridis, C., and Van Gool, L. Advances in deep concealed scene understanding. *Visual Intell.*, 1(1):16, 2023a. 13

- Fan, K., Wang, C., Wang, Y., Wang, C., Yi, R., and Ma, L. Rfenet: towards reciprocal feature evolution for glass segmentation. In *IJCAI*, pp. 717–725, 2023b. 5
- Fang, C., He, C., Tang, L., Zhang, Y., Zhu, C., Shen, Y., Chen, C., Xu, G., and Li, X. Integrating extra modality helps segmentor find camouflaged objects well. *arXiv preprint arXiv:2502.14471*, 2025. 7
- Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., and Torr, P. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(2):652–662, 2019. 6
- Gao, Z., Xu, K., Zhang, X., Zhuang, H., Wan, T., Ding, B., Mao, X., and Huaimin, W. Rethinking obscured sub-optimality in analytic learning for exemplar-free class-incremental learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(10):1123–1136, 2025. doi: 10.1109/TCSVT.2025.3631593. 1
- Han, D., Lee, S., Zhang, C., Yoon, H., Kwon, H., Kim, H.-C., and Choo, H.-G. Internal-external boundary attention fusion for glass surface segmentation. *arXiv preprint arXiv:2307.00212*, 2024. 5
- He, C., Li, K., Zhang, Y., Tang, L., and Zhang, Y. Camouflaged object detection with feature decomposition and edge reconstruction. In *CVPR*, pp. 22046–22055, 2023a. 2, 4, 5
- He, C., Li, K., Zhang, Y., Xu, G., and Tang, L. Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. *NeurIPS*, 2024a. 4, 7
- He, C., Li, K., Zhang, Y., Zhang, Y., Guo, Z., and Li, X. Strategic preys make acute predators: Enhancing camouflaged object detectors by generating camouflaged objects. *ICLR*, 2024b. 1
- He, C., Li, K., Zhang, Y., Yang, Z., Tang, L., Zhang, Y., Kong, L., and Farsiu, S. Segment concealed object with incomplete supervision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025a. 4, 7
- He, C., Shen, Y., Fang, C., Xiao, F., Tang, L., Zhang, Y., Zuo, W., Guo, Z., and Li, X. Diffusion models in low-level vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025b. 1
- He, C., Xiao, F., Zhang, R., Fang, C., Fan, D.-P., and Farsiu, S. Reversible unfolding network for concealed visual perception with generative refinement. *arXiv preprint arXiv:2508.15027*, 2025c. 1
- He, C., Zhang, R., Tang, L., Yang, Z., Li, K., Fan, D.-P., and Farsiu, S. Scaler: Sam-enhanced collaborative learning for label-deficient concealed object segmentation. *arXiv preprint arXiv:2511.18136*, 2025d. 2, 4
- He, C., Zhang, R., Xiao, F., Fang, C., Tang, L., Zhang, Y., Kong, L., Fan, D.-P., Li, K., and Farsiu, S. Run: Reversible unfolding network for concealed object segmentation. *ICML*, 2025e. 1, 5, 6, 7
- He, C., Zhang, R., Zhang, D., Xiao, F., Fan, D.-P., and Farsiu, S. Nested unfolding network for real-world concealed object segmentation. *arXiv preprint arXiv:2511.18164*, 2025f. 2, 8
- He, C., Zhang, R., Zhang, D., Fang, C., Tang, L., Feng, J., Xiao, F., and Farsiu, S. Ride: Retinex-informed decoupling for exposing concealed objects. *arXiv preprint arXiv:2605.15450*, 2026. 1
- He, H., Li, X., Cheng, G., Shi, J., Tong, Y., Meng, G., Prinnet, V., and Weng, L. Enhanced boundary learning for glass-like object segmentation. In *ICCV*, pp. 15859–15868, 2021. 5
- He, J., Li, P., Geng, Y., and Xie, X. Fastinst: A simple query-based model for real-time instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 23663–23672, 2023b. 8
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *CVPR*, pp. 770–778, 2016. 6
- Hu, X., Wang, S., Qin, X., Dai, H., Ren, W., Luo, D., Tai, Y., and Shao, L. High-resolution iterative feedback network for camouflaged object detection. In *AAAI*, volume 37, pp. 881–889, 2023. 6
- Hu, Y., Goel, K., Killiakov, V., and Yang, Y. Eigenspectrum analysis of neural networks without aspect ratio bias. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 24290–24313. PMLR, 13–19 Jul 2025. URL <https://proceedings.mlr.press/v267/hu25e.html>. 1
- Hu, Y., Wang, Y., and McAuley, J. Evaluating memory in LLM agents via incremental multi-turn interactions. In *The Fourteenth International Conference on Learning Representations*, 2026. URL <https://openreview.net/forum?id=DT7JyQC3MR>. 1
- Jain, S., Atale, R., Gupta, A., Mishra, U., Seal, A., Ojha, A., Jaworek-Korjakowska, J., and Krejcar, O. Coinnet: A convolution-involution network with a novel statistical attention for automatic polyp segmentation. *IEEE Trans. Med. Imaging*, 42(12):3987–4000, 2023. 5

- Kumar, M., Packer, B., and Koller, D. Self-paced learning for latent variable models. *Advances in neural information processing systems*, 23, 2010. 2, 7
- Le, T.-N., Nguyen, T. V., Nie, Z., Tran, M.-T., and Sugimoto, A. Anabranched network for camouflaged object segmentation. *Comput. Vis. Image Underst.*, 184:45–56, 2019. 13
- Li, H. and Fu, Y. Fcdfusion: A fast, low color deviation method for fusing visible and infrared image pairs. *Computational Visual Media*, 11(1):195–211, 2025. 2
- Lin, J. and He, Z. Rich context aggregation with reflection prior for glass surface detection. In *CVPR*, pp. 13415–13424, 2021. 13
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. Microsoft coco: Common objects in context. In *European conference on computer vision*, pp. 740–755. Springer, 2014. 8
- Lu, S., Wang, Z., Li, L., Liu, Y., and Kong, A. W.-K. Mace: Mass concept erasure in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6430–6440, 2024. 1
- Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., and Fan, D.-P. Simultaneously localize, segment and rank the camouflaged objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11591–11601, 2021. 13
- Maaten, L. v. d. and Hinton, G. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov): 2579–2605, 2008. 8
- Margolin, R., Zelnik-Manor, L., and Tal, A. How to evaluate foreground maps? In *CVPR*, pp. 248–255, 2014. 13
- Matiisen, T., Oliver, A., Cohen, T., and Schulman, J. Teacher–student curriculum learning. *IEEE transactions on neural networks and learning systems*, 31(9):3732–3740, 2019. 2, 7
- Mei, H., Yang, X., Wang, Y., Liu, Y., and He, S. Don’t hit me! glass detection in real-world scenes. In *CVPR*, pp. 3687–3696, 2020. 1, 13
- Mei, H., Yang, X., Yu, L., Zhang, Q., Wei, X., and Lau, R. W. Large-field contextual feature learning for glass detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023. 13
- Ni, Z., Chen, X., Zhai, Y., Tang, Y., and Wang, Y. Context-guided spatial feature reconstruction for efficient semantic segmentation. In *European Conference on Computer Vision*, pp. 239–255. Springer, 2024. 8
- Pang, Y., Zhao, X., Xiang, T.-Z., Zhang, L., and Lu, H. Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In *CVPR*, pp. 2160–2170, 2022. 1, 2, 8
- Pang, Y., Zhao, X., Xiang, T.-Z., Zhang, L., and Lu, H. Zoomnext: A unified collaborative pyramid network for camouflaged object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024. 2, 7
- Pentina, A., Sharmanska, V., and Lampert, C. H. Curriculum learning of multiple tasks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5492–5500, 2015. 2
- Saglietti, L., Mannelli, S., and Saxe, A. An analytical theory of curriculum learning in teacher-student networks. *Advances in Neural Information Processing Systems*, 35: 21113–21127, 2022. 2
- Shen, Y., Xiao, F., Hu, S., Pang, Y., Pu, Y., Fang, C., Li, X., and He, C. Uncertainty-masked bernoulli diffusion for camouflaged object detection refinement. *arXiv preprint arXiv:2506.10712*, 2025. 2
- Silva, J., Histace, A., Romain, O., and Dray, X. Toward embedded detection of polyps in wce images for early diagnosis. *Int. J. Comput. Assist. Radiol. Surg.*, 9:283–293, 2014. 13
- Skurowski, P., Abdulameer, H., and Błaszczuk, J. Animal camouflage analysis: Chameleon database. *Unpublished manuscript*, pp. 7, 2018. 13
- Soviany, P. Curriculum learning with diversity for supervised computer vision tasks. *arXiv preprint arXiv:2009.10625*, 2020. 2
- Soviany, P., Ionescu, R. T., Rota, P., and Sebe, N. Curriculum learning: A survey. *International Journal of Computer Vision*, 130(6):1526–1565, 2022. 2
- Sun, J., Xu, K., Pang, Y., Zhang, L., Lu, H., Hancke, G., and Lau, R. Adaptive illumination mapping for shadow detection in raw images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12709–12718, 2023. 8
- Sun, K., Chen, Z., Lin, X., Sun, X., Liu, H., and Ji, R. Conditional diffusion models for camouflaged and salient object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 5
- Sun, Y., Xu, C., Yang, J., Xuan, H., and Luo, L. Frequency-spatial entanglement learning for camouflaged object detection. In *ECCV*, pp. 343–360, 2024. 2, 5, 8

- Tajbakhsh, N., Gurudu, S. R., and Liang, J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Trans. Med. Imaging*, 35(2): 630–644, 2015. 1, 13
- Tullis, J. G. and Benjamin, A. S. On the effectiveness of self-paced learning. *Journal of memory and language*, 64(2):109–118, 2011. 2
- Vicente, T. F. Y., Hoai, M., and Samaras, D. Leave-one-out kernel optimization for shadow detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3388–3396, 2015. 8
- Vicente, T. F. Y., Hou, L., Yu, C.-P., Hoai, M., and Samaras, D. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *European conference on computer vision*, pp. 816–832. Springer, 2016. 8
- Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., and Shao, L. Pvt v2: Improved baselines with pyramid vision transformer. *Comput. Vis. Media*, 8(3):415–424, 2022. 6
- Wang, W., Sun, H., and Wang, X. Lssnet: A method for colon polyp segmentation based on local feature supplementation and shallow feature supplementation. In *MICCAI*, pp. 446–456. Springer, 2024. 5, 13
- Xiao, F., Hu, S., Shen, Y., and He, C. A survey of camouflaged object detection and beyond. *arXiv preprint arXiv:2408.14562*, 2024. 1
- Xiao, F., Feng, J., Hu, P., Zhang, D., Xu, L., Qin, G., Li, L., He, C., and Farsiu, S. Qualiteacher: Quality-conditioned pseudo-labeling for real-world image restoration. *arXiv preprint arXiv:2603.08030*, 2026a. 1
- Xiao, F., Hu, P., Xu, L., Guo, X., Qin, G., Shen, Y., Fang, C., Zhang, R., He, C., and Farsiu, S. Beyond ground-truth: Leveraging image quality priors for real-world image restoration. *arXiv preprint arXiv:2603.29773*, 2026b. 1
- Yan, T., Gao, J., Xu, K., Zhu, X., Huang, H., Li, H., Wah, B., and Lau, R. W. Ghostingnet: A novel approach for glass surface detection with ghosting cues. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024. 5
- Yang, F., Zhai, Q., Li, X., Huang, R., Luo, A., Cheng, H., and Fan, D.-P. Uncertainty-guided transformer reasoning for camouflaged object detection. In *ICCV*, pp. 4146–4155, 2021. 8, 13
- Yang, X., Zhu, H., Mao, G., and Xing, S. Oaformer: Occlusion aware transformer for camouflaged object detection. In *ICME*, pp. 1421–1426. IEEE, 2023. 6
- Yin, B., Zhang, X., Fan, D.-P., Jiao, S., Cheng, M.-M., Van Gool, L., and Hou, Q. Camoformer: Masked separable attention for camouflaged object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024. 6
- Yu, Z., Zhang, X., Zhao, L., Bin, Y., and Xiao, G. Exploring deeper! segment anything model with depth perception for camouflaged object detection. In *ACM MM*, pp. 4322–4330, 2024. 7
- Zhai, Q., Li, X., Yang, F., and Chen, C. Mutual graph learning for camouflaged object detection. In *CVPR*, pp. 2997–3007, 2021. 5
- Zhang, M., Xu, S., Piao, Y., Shi, D., Lin, S., and Lu, H. Preynet: Preying on camouflaged objects. In *ACM MM*, pp. 5323–5332, 2022a. 5
- Zhang, M., Zhang, R., Yang, Y., Bai, H., Zhang, J., and Guo, J. Isnet: Shape matters for infrared small target detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 877–886, 2022b. 8
- Zhang, M., Wang, Y., Guo, J., Li, Y., Gao, X., and Zhang, J. Irsam: Advancing segment anything model for infrared small target detection. In *European Conference on Computer Vision*, pp. 233–249. Springer, 2024. 8
- Zhang, X., Yin, B., Lin, Z., Hou, Q., Fan, D.-P., and Cheng, M.-M. Referring camouflaged object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2
- Zhao, X., Pang, Y., Ji, W., Sheng, B., Zuo, J., Zhang, L., and Lu, H. Spider: a unified framework for context-dependent concept segmentation. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 60906–60926, 2024. 1, 8
- Zheng, Y., Zhong, B., Liang, Q., Zhang, S., Li, G., Li, X., and Ji, R. Towards universal modal tracking with online dense temporal token learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 1
- Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., and Torralba, A. Scene parsing through ade20k dataset. In *CVPR*, pp. 633–641, 2017. 8
- Zhu, H., Li, P., Xie, H., Yan, X., Liang, D., Chen, D., Wei, M., and Qin, J. I can find you! boundary-guided separated attention network for camouflaged object detection. In *AAAI*, volume 36, pp. 3608–3616, 2022. 5

## A. Experiment

### A.1. Datasets and Metrics

**Camouflaged object detection.** Following the protocol of SINet (Fan et al., 2020), we conduct experiments on four datasets: *CHAMELEON* (Skurowski et al., 2018), *CAMO* (Le et al., 2019), *COD10K* (Fan et al., 2021a), and *NC4K* (Lv et al., 2021). The *CHAMELEON* dataset contains 76 images, while *CAMO* comprises 1,250 images from 8 categories. *COD10K* includes 5,066 images organized into 10 super-classes, and *NC4K* is the largest test set, with 4,121 images. For training, we use 1,000 images from *CAMO* and 3,040 images from *COD10K*; the remaining images from these two datasets, together with all images from *CHAMELEON* and *NC4K*, are reserved for testing. We report four widely adopted metrics: mean absolute error ( $M$ ), adaptive F-measure ( $F_\beta$ ) (Margolin et al., 2014), mean E-measure ( $E_\phi$ ) (Fan et al., 2021b), and structure measure ( $S_\alpha$ ) (Fan et al., 2017). Lower  $M$  and higher  $F_\beta$ ,  $E_\phi$ , and  $S_\alpha$  indicate superior performance.

**Polyp image segmentation.** For polyp segmentation, we evaluate on two public benchmarks: *CVC-ColonDB* (Tajbakhsh et al., 2015) and *ETIS* (Silva et al., 2014). The training protocol follows the setting of LSSNet (Wang et al., 2024). Quantitative performance is assessed using three commonly employed metrics: mean Dice (mDice), mean Intersection over Union (mIoU), and structure measure ( $S_\alpha$ ), where larger values indicate better performance. To ensure a fair comparison with recent state-of-the-art approaches, which predominantly adopt transformer-based encoders, we use PVT V2 as the encoder backbone.

**Transparent object detection.** For transparent object detection, we also adopt PVT V2 as the default backbone to enable a fair comparison with existing methods. Experiments are conducted on two datasets, *GDD* (Mei et al., 2020) and *GSD* (Lin & He, 2021). The training set consists of 2,980 images from *GDD* and 3,202 images from *GSD*, with the remaining images used exclusively for inference. Following GNet-B (Mei et al., 2023), we evaluate performance using mIoU and the maximum F-measure ( $F_\beta^{max}$ ), together with mean absolute error ( $M$ ). Better performance corresponds to lower  $M$  and higher mIoU and  $F_\beta^{max}$ .

**Concealed defect detection.** For concealed defect detection, we employ PVT V2 as the default backbone and assess the generalization capability of the proposed RUN framework. Specifically, we directly apply the model trained on the COD task to segment concealed defects on the *CDS2K* dataset (Fan et al., 2023a). Five evaluation metrics are reported; for all metrics except MAE, larger values indicate better performance, whereas for MAE lower values are preferable.

### A.2. Further Analysis

#### A.2.1. IMPLEMENTATION DETAILS OF “COMPATIBILITY WITH ADVANCED ARCHITECTURES”

“\*” in Table 11 indicates simple architecture-specific adaptations, bringing better performance. The implementation details are presented as follows:

**Curri\* (Attention):** During SBFT, we randomly mask a frequency band within radius range  $[r_{low}, r_{high}]$  where  $r_{high} - r_{low} = 0.05$ , encouraging the attention mechanism to discover scale-invariant patterns.

**Curri\* (Multi-scale):** We apply multi-scale SBFT with different cutoff radii  $r$  per scale, and tighten the uncertainty thresholds  $W_{min}$  and  $W_{min}^s$  in RCS, as multi-scale features inherently reduce prediction instability.

**Curri\* (Uncertainty):** We combine the feature-level uncertainty proposed by UGTR (Yang et al., 2021) from temporal statistics with the existing pixel-level uncertainty, jointly suppressing unreliable gradients at both granularities.

The consistent improvements demonstrate that CurriSeg complements rather than conflicts with existing design principles, offering a promising direction for enhancing diverse segmentation frameworks.