EReCu: Pseudo-label Evolution Fusion and Refinement with Multi-Cue Learning for Unsupervised Camouflage Detection

Supplementary Material

1. Hyperparameter Experiments

Hyperparameter selection plays a pivotal role in model performance. This section presents systematic ablation studies on key hyperparameters of the framework's core modules, examining their effects on model performance and substantiating the rationality of the selected optimal configurations. All experiments are conducted on the CAMO-Test and COD10K-Test benchmark datasets, employing Structure Measure (S_m), Weighted F-measure (F_ω^β), and Mean Absolute Error (M) as evaluation metrics. At the same time, all remaining parameters are fixed to their optimal settings as detailed in the main paper.

1.1. Hyperparameters of the MNP Module

The key hyperparameters of the Multi-Cue Native Perception (MNP) module are the number of sampling iterations, N, and the sampling patch size, K. These parameters affect both the stability of the multi-cue quality metric S_{mc} and the model's ability to discriminate boundaries.

1.1.1. Number of Random Sampling Rounds N

In the cosine similarity computation of Eq. (5), the parameter N denotes the number of random sampling iterations applied to $K \times K$ patches, which mitigates estimation bias in similarity metrics caused by irregular region boundaries. We conduct systematic evaluations across $N \in \{1, 3, 5, 7, 9\}$, with results presented in Tab. 3.

Optimal performance is achieved when $N = 5$. For $N < 5$, the limited number of sampling iterations yields inadequate similarity estimation and imprecise boundary discrimination. In contrast, when $N > 5$, performance deteriorates monotonically, as the increased computational cost of additional sampling rounds does not translate into enhanced robustness of the derived metrics.

1.1.2. Size of Sampling Patches K

The parameter $K \times K$ characterizes the spatial resolution of sampling patches, which directly influences the granularity at which local texture and semantic features are captured. Systematic experiments evaluate performance variations across $K \in \{5, 10, 15, 20, 25\}$, with corresponding results summarized in Tab. 4.

Optimal performance is attained when $K = 15$. A patch size of $K = 5$ is insufficient to capture complete local texture features, while $K > 20$ results in excessive coverage that introduces background noise and degrades foreground-background discrimination.

1.2. Hyperparameter Collaboration Analysis

The synergistic collaboration between hyperparameters N and K in the MNP module is visualized through a heatmap in Fig. 7. The results demonstrate that the combination $N = 5$ and $K = 15$ achieves optimal performance with strong positive synergy, thereby validating the rationality of our hyperparameter selections.

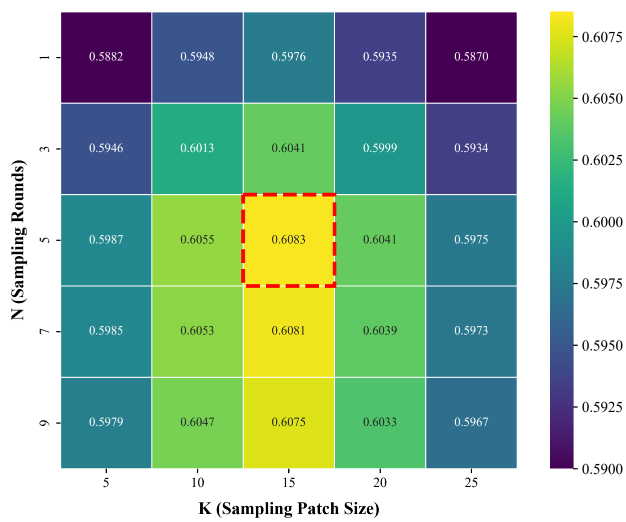

Figure 7. Heatmap of the impact of N and K collaboration on F_ω^β

2. Supplementary Ablation Studies

2.1. Comparison of Texture Extractors

The MNP module leverages a synergistic combination of Local Binary Pattern (LBP) and Difference of Gaussians (DoG) to extract low-level texture representations. Comparative ablation studies across various texture extractors are presented in Table Tab. 5. The LBP+DoG combination demonstrates complementary strengths in capturing both texture edges and scale variations, consistently outperforming individual extractor approaches.

For a 3×3 neighborhood centered at pixel (x_c, y_c), the LBP descriptor is computed by thresholding each of the eight neighboring pixels against the central pixel value and concatenating the results into a binary string:

$$\text{LBP}(x_c, y_c) = \sum_{p=1}^{8} s(I(x_p, y_p) - I(x_c, y_c)) \cdot 2^{p-1} \quad (16)$$

Table 3. Impact of different N values on model performance

Dataset	Metric	w/o MNP	$N=1$	$N=3$	$N=5$	$N=7$	$N=9$
CAMO-Test	$S_m \uparrow$	.6887	.6942	.6998	.7027	**.7029**	.7021
	$F_\omega^\beta \uparrow$	.5923	.5976	.6041	**.6083**	.6081	.6075
	$M \downarrow$	.1182	.1125	.1088	**.1072**	.1075	.1082
COD10K-Test	$S_m \uparrow$	.7111	.7154	.7196	.7221	**.7224**	.7217
	$F_\omega^\beta \uparrow$	.5478	.5529	.5587	**.5628**	.5625	.5618
	$M \downarrow$	.0653	.0641	.0627	**.0613**	.0615	.0620

Table 4. Impact of different K values on model performance

Dataset	Metric	w/o MNP	$K=5$	$K=10$	$K=15$	$K=20$	$K=25$
CAMO-Test	$S_m \uparrow$	.6887	.6952	.7018	**.7027**	.6996	.6931
	$F_\omega^\beta \uparrow$	.5923	.5987	.6055	**.6083**	.6041	.5975
	$M \downarrow$	.1182	.1121	.1083	**.1072**	.1094	.1127
COD10K-Test	$S_m \uparrow$	.7111	.7165	.7208	**.7221**	.7189	.7124
	$F_\omega^\beta \uparrow$	.5478	.5536	.5594	**.5628**	.5589	.5521
	$M \downarrow$	.0653	.0637	.0621	**.0613**	.0623	.0638

where the sign function $s(\cdot)$ is defined as:

$$s(z) = \begin{cases} 1 & \text{if } z \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

where $I(x_c, y_c)$ denotes the grayscale intensity at the central pixel location, $I(x_p, y_p)$ represents the grayscale intensity of the p-th neighboring pixel (enumerated clockwise starting from the top-left position), with $p \in \{1, 2, \ldots, 8\}$ indexing the complete set of eight surrounding pixels.

For two Gaussian kernels characterized by distinct standard deviations σ_1 and σ_2 with $\sigma_2 > \sigma_1$, DoG operator at pixel location (x, y) is formally defined as:

$$\begin{aligned} \text{DoG}(x, y, \sigma_1, \sigma_2) = & \frac{1}{2\pi\sigma_2^2} \exp\left(-\frac{x^2+y^2}{2\sigma_2^2}\right) \\ & - \frac{1}{2\pi\sigma_1^2} \exp\left(-\frac{x^2+y^2}{2\sigma_1^2}\right) \end{aligned} \tag{18}$$

2.2. Comparison of Semantic Extractors

Comparative ablation experiments on various semantic feature extractors are provided in Tab. 6. While MobileNetV2 and EfficientNet-B0 demonstrate limited performance gains, ResNet-18 offers a favorable balance between discriminative capacity and computational efficiency.

2.3. Visualization of MNP Features

As illustrated in Fig. 8, the MNP module extracts texture features (LBP+DoG), semantic features (hierarchical outputs from ResNet-18), and their fused multi-cue representation. The visualization reveals that the fused features effectively retain both low-level edge and texture characteristics of the target object and capture high-level semantic context, collectively furnishing informative perceptual cues essential for pseudo-label evolution.

Table 5. Performance comparison of different texture extractors on CAMO and COD10K datasets.

Texture Extractor	CAMO-Test			COD10K-Test		
	$S_m \uparrow$	$F_\omega^\beta \uparrow$	$M \downarrow$	$S_m \uparrow$	$F_\omega^\beta \uparrow$	$M \downarrow$
LBP	.6958	.6031	.1148	.7173	.5586	.0637
DoG	.7013	.6058	.1104	.7202	.5607	.0625
LBP + DoG	**.7027**	**.6083**	**.1072**	**.7221**	**.5628**	**.0613**
w/o MNP	.6887	.5923	.1182	.7111	.5478	.0653

Table 6. Performance comparison of different semantic extractors on CAMO and COD10K datasets.

Semantic Extractor	CAMO-Test			COD10K-Test		
	$S_m \uparrow$	$F_\omega^\beta \uparrow$	$M \downarrow$	$S_m \uparrow$	$F_\omega^\beta \uparrow$	$M \downarrow$
MobileNetV2	.6934	.6048	.1105	.7156	.5554	.0631
EfficientNet-B0	.7017	.6067	.1095	.7209	.5613	.0624
ResNet-18	**.7027**	**.6083**	**.1072**	**.7221**	**.5628**	**.0613**
w/o MNP	.6887	.5923	.1182	.7111	.5478	.0653

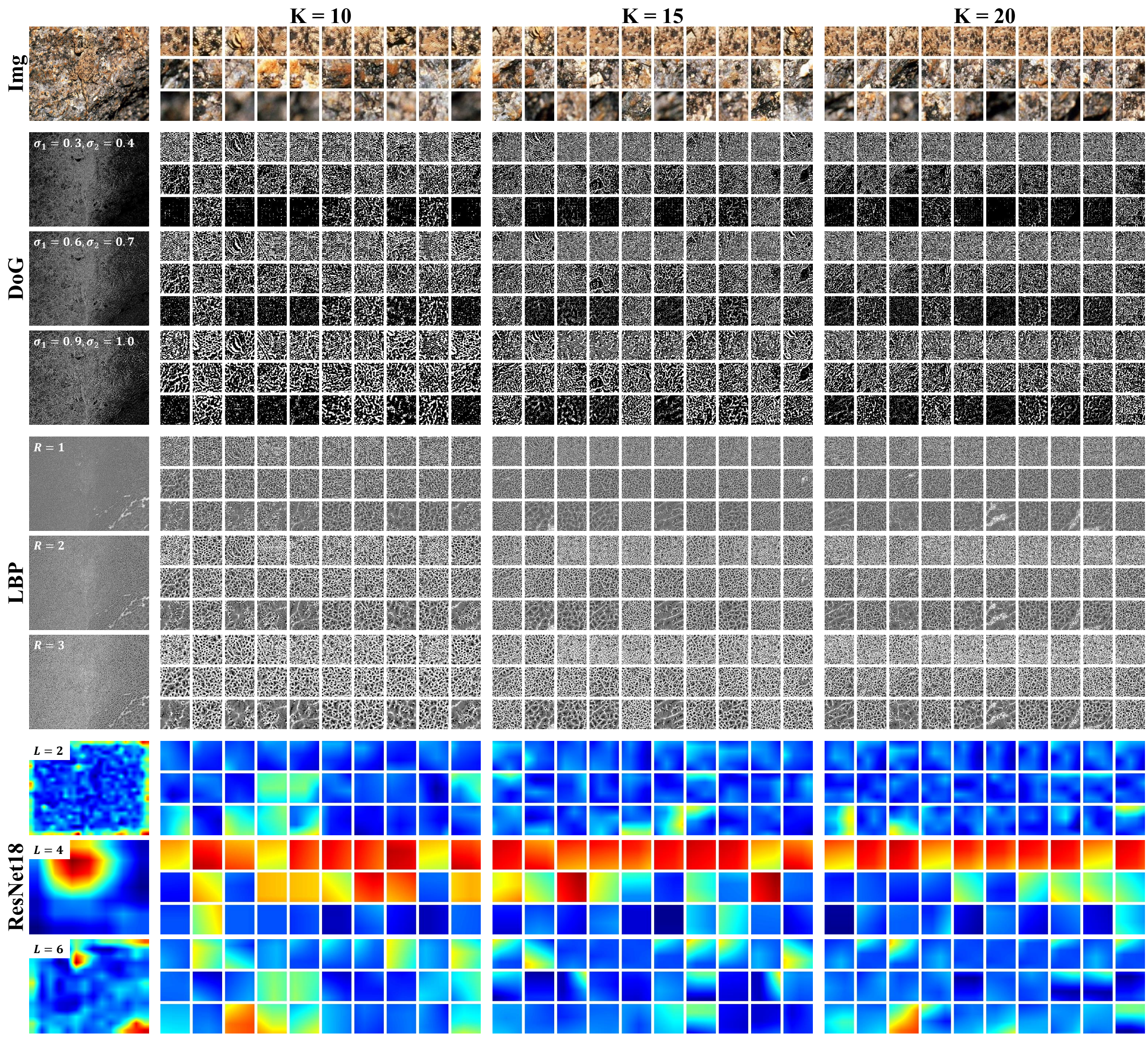

Figure 8. Visualization of different feature extractors under various parameters in the MNP module. Each group consists of three rows, each containing 10 $K \times K$ patches. Specifically, the first row shows randomly selected patches within the Mask, the second row shows randomly selected patches around the Mask, and the third row shows randomly selected patches outside the Mask.