FCL-COD: Weakly Supervised Camouflaged Object Detection with Frequency-aware and Contrastive Learning

Supplementary Material

This supplementary material presents: (1) extended comparisons with state-of-the-art methods; (2) in-depth ablation analyses of FoRA, GCL, and MSFA modules; (3) hyperparameter sensitivity studies; and (4) generalization experiments on Salient Object Detection.

1. Comprehensive Comparison with State-of-the-art Methods

For completeness and transparency, we present the full, unabridged quantitative comparison table in Table 1. While the main manuscript provides a summarized version due to space limitations, this table offers a comprehensive overview of our method's performance against competing approaches across all four benchmark datasets, including CAMO, COD10K, NC4K, and CHAMELEON. The detailed results further affirm the conclusions drawn in the main paper, consistently demonstrating the superior performance of our FCL-COD model over all other state-of-the-art weakly supervised methods across all evaluated metrics.

In addition to the quantitative results, we provide more extensive qualitative comparisons to further substantiate the superiority and robustness of our proposed method. Figure 2 presents a detailed side-by-side comparison of our method against eight recent state-of-the-art camouflaged object detection methods: CamoFormer, CRNet, FSPNet, SARNet, UGTR, WS-SAM, and ZoomNet. The selected scenes feature a variety of challenging conditions, including objects with extremely high texture similarity to the background, small object sizes, and intricate boundaries.

As is visually evident from the comparison, our method consistently yields prediction masks that are remarkably closer to the Ground Truth. These extended qualitative results supplement the quantitative analysis in the main paper, offering compelling visual evidence that our method achieves state-of-the-art performance. It demonstrates a superior ability to maintain object integrity, accurately delineate boundaries, and minimize false positives, handling diverse and challenging camouflage scenarios more effectively than existing approaches.

2. Extended Qualitative Ablation Studies

To provide a comprehensive qualitative evaluation corresponding to the ablation experiments in the main manuscript, we present extended visual analyses for each key module: the Frequency-aware Low-Rank Adaptation (FoRA), the Gradient-aware Contrastive Learning (GCL), the Multi-Scale Frequency-aware Attention (MSFA), and the overall progressive integration.

Algorithm 1 Spectral Energy Analysis Pipeline

Input: Feature map $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$
Output: Differential Energy Distribution $\Delta E(r)$

Compute frequency spectrum: $\mathcal{F} \leftarrow \text{FFT2D}(\mathbf{F})$
Calculate spectral energy density: $S(u,v) \leftarrow \sum_c |\mathcal{F}_c(u,v)|^2$
Compute normalized radial frequency: $r(u,v) \in [0,1]$
Partition spectrum into bands based on r:
  $\mathcal{B}_{\text{Low}} : 0 \le r \le 0.25$
  $\mathcal{B}_{\text{Mid}} : 0.25 < r \le 0.5$
  $\mathcal{B}_{\text{High}} : r > 0.5$
Calculate energy for FoRA and baseline LoRA within each band
Compute differential energy: $\Delta E(r) \leftarrow E_{\text{FoRA}}(r) - E_{\text{LoRA}}(r)$
return $\Delta E(r)$

2.1. Extended Qualitative Analysis of FoRA

To provide deeper insights into FoRA's efficacy, we perform a spectral energy analysis as detailed in Algorithm 1. For a given feature map $\mathbf{F}$, we compute its frequency spectrum $\mathcal{F}$ via 2D FFT and partition the domain into three bands based on the normalized radial frequency $r \in [0,1]$: **Low Frequency** ($0 \le r \le 0.25$), **Mid Frequency** ($0.25 < r \le 0.5$), and **High Frequency** ($r > 0.5$). The differential energy $\Delta E(r) = E_{\text{FoRA}}(r) - E_{\text{LoRA}}(r)$, plotted in Figures 3 and 4, reveals the frequency-selective mechanism.

As indicated by the "Frequency Change Curves", FoRA exhibits a distinct strategy: it selectively suppresses low-frequency components (red regions, $r < 0.2$) while significantly boosting mid-frequency energy (green peaks, $0.2 \le r \le 0.6$). This mechanism effectively filters out background noise—which often shares low-frequency statistics with the foreground—while amplifying structural cues essential for object integrity. Consequently, the visual comparisons show that FoRA successfully connects fragmented features and fills semantic voids, producing coherent segmentation masks dominated by green True Positive pixels, whereas the baseline LoRA yields incomplete predictions with extensive red False Negative regions.

2.2. Effectiveness of the GCL Module with Grad-CAM

The qualitative results in Figure 5 further illustrate the advantage of our GCL module. Grad-CAM visualizations reveal that our **GCL** module compels the model to focus on background regions that are texturally and chromatically similar to the foreground object. These regions represent hard negative samples, which are crucial for effective contrastive learning. In contrast, the model without GCL exhibits diffuse and unfocused attention, often mistaking confusing background patches as part of the target. By explicitly mining these hard negatives through gradient-aware contrast, GCL enforces a larger margin between foreground and confusing background features in the embedding space, significantly reducing false positives in challenging camouflaged scenes where traditional pixel-level supervision fails. This targeted focus empowers our model to better distinguish the object from its challenging surroundings, leading to more precise segmentation masks with fewer spurious activations.

2.3. Extended Qualitative Analysis of MSFA

We further visualize the ablation of the Multi-Scale Frequency-aware Attention (MSFA) module in Figure 6. The visualization reveals progressive quality improvements: the **Base** model produces coarse boundaries and incomplete objects. Adding Tri-Channel Attention ($\mathcal{T}$) enables cross-scale gating but remains single-domain. Introducing the spatial branch ($\mathcal{T} + \mathcal{M}_{spa}$) enhances localization by capturing local context through stacked convolutions, reducing false positives. The frequency branch ($\mathcal{T}+\mathcal{M}_{fre}$) further refines boundary sharpness by modeling spectral cues in the Fourier domain. Finally, our full MSFA (**Ours**) integrates both branches across three scales—small, medium, and large—suppressing background noise while preserving fine structures like insect legs. This confirms that multi-scale spatial-frequency interaction is essential: spatial context aids localization, frequency cues enhance boundary delineation, and cross-scale fusion ensures fine-grained details are preserved in deep semantic layers—particularly vital for detecting small or thin camouflaged targets.

2.4. Overall Progressive Integration

Figure 7 provides a holistic qualitative comparison that complements our quantitative results, demonstrating the step-by-step efficacy of our key components. As illustrated, the **Baseline** model yields only coarse predictions characterized by incomplete structures and imprecise boundaries. The integration of **FoRA** significantly improves core region identification, resulting in more complete initial segmentation maps. Subsequently, adding the **GCL** module enhances global context understanding, which facilitates better object-background distinction and shape refinement. Finally, our **Full Model**, equipped with the MSFA module, achieves the most accurate segmentation by leveraging multi-scale feature fusion and edge refinement capabilities. These qualitative results corroborate our quantitative findings, validating the essential contribution of each module to the overall network performance.

3. Analysis of Hyperparameters in Anchor Loss

We further conduct experiments to analyze the sensitivity of our model to the hyperparameters in the anchor loss, $\mathcal{L}_{\text{anchor}}$. As defined in the main manuscript, this loss is controlled by two weighting parameters, $\lambda_{\text{stu}}^{\text{dice}}$ and $\lambda_{\text{tea}}^{\text{dice}}$, which balance the consistency constraints between the self-training model and the frozen anchor model. To investigate their impact, we vary the values of both $\lambda_{\text{stu}}^{\text{dice}}$ and $\lambda_{\text{tea}}^{\text{dice}}$ within the range of [0, 1] and record the corresponding model performance. The results are visualized in Figure 1. As shown in the line graphs, the model's performance peaks when both $\lambda_{\text{stu}}^{\text{dice}}$ and $\lambda_{\text{tea}}^{\text{dice}}$ are set to 0.5. This indicates that an equal balance between the two consistency terms yields the optimal result. Based on this analysis, we set $\lambda_{\text{stu}}^{\text{dice}} = 0.5$ and $\lambda_{\text{tea}}^{\text{dice}} = 0.5$ for all our experiments.

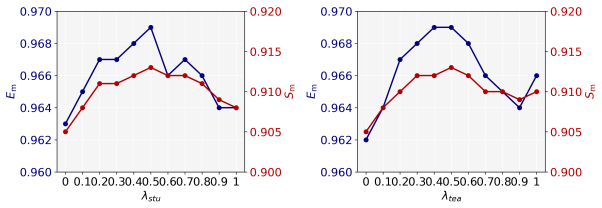

Figure 1. Hyperparameter Sensitivity Analysis of $\lambda_{\text{stu}}^{\text{dice}}$ and $\lambda_{\text{tea}}^{\text{dice}}$ in Anchor Loss

4. Generalization Performance on Salient Object Detection

To further demonstrate the generalization capability of our model, we provide extended results for the Salient Object Detection task across four standard benchmarks. As shown in Table 2, our FCL-COD achieves the best performance on all three metrics for ECSSD and DUT-O, and obtains the best MAE of 0.020 and F-measure of 0.954 on HKU-IS, surpassing most fully-supervised methods despite using only box-level supervision. Notably, while SOD targets visually prominent objects with distinct appearance characteristics from camouflaged scenarios, our frequency-aware modules and contrastive learning mechanisms remain highly effective by capturing generalizable boundary and context cues. This cross-task effectiveness validates that our architectural design successfully learns transferable visual representations beyond task-specific biases.

Table 1. Comprehensive evaluation of FCL-COD against competing methods across four benchmarks

Methods	Sup.	CAMO				COD10K				NC4K				CHAMELEON			
		MAE↓	S_m↑	E_m↑	F_β^w↑	MAE↓	S_m↑	E_m↑	F_β^w↑	MAE↓	S_m↑	E_m↑	F_β^w↑	MAE↓	S_m↑	E_m↑	F_β^w↑
Fully-Supervised Methods (F)																	
SINet [3]	F	0.092	0.745	0.804	0.644	0.043	0.776	0.864	0.631	0.058	0.808	0.871	0.723	0.034	0.872	0.946	0.806
MGL-R [23]	F	0.088	0.775	0.812	0.673	0.035	0.814	0.851	0.666	0.052	0.833	0.867	0.740	0.030	0.893	0.941	0.812
PFNet [15]	F	0.085	0.782	0.841	0.695	0.040	0.800	0.877	0.660	0.053	0.829	0.887	0.745	0.033	0.882	0.945	0.810
UGTR [20]	F	0.086	0.784	0.822	0.684	0.036	0.817	0.852	0.666	0.052	0.839	0.874	0.747	0.030	0.891	0.955	0.833
UJSC [12]	F	0.073	0.800	0.859	0.728	0.035	0.809	0.884	0.684	0.047	0.842	0.898	0.771	0.030	0.891	0.955	0.833
ZoomNet [16]	F	0.066	0.820	0.892	0.752	0.029	0.838	0.911	0.729	0.043	0.853	0.896	0.784	0.023	0.902	0.958	0.845
SAM-Adapter [2]	F	0.070	0.847	0.873	0.765	0.025	0.883	0.918	0.801	-	-	-	-	0.033	0.896	0.919	0.824
C2FNet [17]	F	0.080	0.796	0.864	0.719	0.036	0.813	0.900	0.686	0.049	0.838	0.904	0.762	0.032	0.888	0.946	0.828
SegMaR [10]	F	0.071	0.815	0.884	0.753	0.034	0.833	0.906	0.724	0.046	0.841	0.907	0.781	0.025	0.906	0.959	0.860
CamoFormer-R [21]	F	0.067	0.817	0.885	0.752	0.029	0.838	0.930	0.724	0.042	0.855	0.914	0.788	0.025	0.898	0.956	0.847
DGNet [9]	F	0.057	0.838	0.914	0.768	0.033	0.822	0.911	0.758	0.042	0.857	0.922	0.783	0.029	0.890	0.956	0.816
SLSR [14]	F	0.080	0.787	0.854	0.696	0.037	0.804	0.892	0.673	0.048	0.840	0.907	0.765	0.030	0.890	0.948	0.822
BSA-Net [24]	F	0.079	0.794	0.867	0.717	0.034	0.818	0.901	0.699	0.048	0.842	0.907	0.771	0.027	0.895	0.957	0.841
BGNet [18]	F	0.073	0.812	0.882	0.749	0.033	0.831	0.911	0.722	0.044	0.851	0.916	0.788	0.027	0.901	0.954	0.851
FEDER [4]	F	0.071	0.802	0.873	0.738	0.032	0.822	0.905	0.716	0.044	0.847	0.915	0.789	0.030	0.887	0.954	0.835
CamoFormer-P [21]	F	**0.046**	0.872	**0.938**	0.831	0.023	0.869	0.939	0.786	**0.030**	**0.892**	**0.946**	0.847	0.022	0.910	0.966	0.865
MSCAF-Net [13]	F	**0.046**	0.873	0.937	0.828	0.024	0.865	0.936	0.775	0.032	0.887	0.942	0.838	0.022	0.912	0.970	0.865
HitNet [7]	F	0.055	0.849	0.910	0.809	0.023	0.871	0.938	0.806	0.037	0.875	0.929	0.834	0.019	0.921	0.972	0.897
FSPNet [8]	F	0.050	0.856	0.928	0.799	0.026	0.851	0.930	0.735	0.035	0.878	0.937	0.816	0.023	0.908	0.965	0.851
SARNet [19]	F	**0.046**	**0.874**	0.935	**0.844**	**0.021**	**0.885**	**0.947**	**0.820**	0.032	0.889	0.940	**0.851**	**0.017**	**0.933**	**0.978**	**0.909**
Weakly Supervised Methods (S, P, B)																	
SAM [11] (SAM-H)	-	0.132	0.684	0.687	0.606	0.050	0.783	0.798	0.701	0.078	0.767	0.776	0.696	0.081	0.727	0.734	0.639
SCSOD [22]	S	0.102	0.713	0.795	0.618	0.055	0.710	0.805	0.546	-	-	-	-	0.053	0.792	0.881	0.714
CRNet [6]	S	0.092	0.735	0.815	0.641	0.049	0.733	0.832	0.576	0.063	0.775	0.855	0.688	0.046	0.818	0.897	0.791
SAM-S [11] (SAM-H)	S	0.105	0.731	0.774	-	0.046	0.772	0.828	-	0.071	0.763	0.832	-	0.076	0.650	0.820	0.729
WS-SAM [5] (SAM-H)	S	0.092	0.759	0.818	-	0.038	0.803	0.878	-	0.052	0.829	0.886	-	0.046	0.824	0.897	0.777
SAM-P [11] (SAM-H)	P	0.123	0.677	0.693	-	0.069	0.765	0.796	-	0.082	0.776	0.786	-	0.101	0.697	0.745	0.696
WS-SAM [5] (SAM-H)	P	0.102	0.718	0.757	-	0.039	0.790	0.856	-	0.057	0.813	0.859	-	0.056	0.805	0.868	0.767
SAM-COD [1](SAM-H)	B	0.062	0.837	0.901	0.786	0.028	0.842	0.914	0.745	0.037	0.867	0.923	0.813	-	-	-	-
FCL-COD(SAM-B)	B	0.060	0.841	0.899	0.795	0.027	0.859	0.924	0.774	0.041	0.867	0.919	0.817	0.039	0.866	0.928	0.799
FCL-COD(SAM-L)	B	0.054	0.856	0.910	0.818	**0.022**	**0.881**	**0.938**	**0.812**	0.034	**0.886**	**0.930**	**0.847**	**0.026**	**0.901**	**0.954**	**0.856**
FCL-COD(SAM-H)	B	**0.050**	**0.862**	**0.915**	**0.824**	**0.022**	0.878	0.934	0.808	**0.033**	0.885	0.928	0.846	0.038	0.882	0.932	0.842

Table 2. Extended Experiments of FCL-COD on Salient Object Detection

	Label	ECSSD			DUT-O			HKU-IS			DUTS-TE		
		MAE↓	S_m↑	F_β^w↑	MAE↓	S_m↑	F_β^w↑	MAE↓	S_m↑	F_β^w↑	MAE↓	S_m↑	F_β^w↑
AFNet	F	0.042	0.913	0.935	0.057	0.826	0.797	0.036	0.905	0.923	0.046	0.867	0.863
BASNet	F	0.037	0.916	0.943	0.057	0.836	0.805	0.032	0.909	0.928	0.048	0.866	0.859
GateNet	F	0.040	0.920	0.945	0.055	0.838	0.818	0.033	0.915	0.934	0.040	0.885	0.888
ICON-R	F	0.032	0.928	0.943	0.057	0.845	0.799	0.029	0.920	0.931	0.037	0.890	0.876
MENet	F	0.031	0.928	0.955	0.045	0.850	0.834	0.023	0.927	0.948	**0.028**	0.905	**0.912**
VST-S++	F	0.027	0.939	0.951	0.050	0.859	0.813	0.025	0.932	0.941	0.029	**0.909**	0.897
SCSOD	S	0.049	0.881	0.914	0.060	0.811	0.782	0.038	0.882	0.908	0.049	0.853	0.858
PSOD	P	0.036	0.913	0.935	0.064	0.824	0.808	0.033	0.901	0.923	0.045	0.853	0.858
SAM-COD	B	0.031	0.929	0.952	0.051	0.844	0.828	0.023	**0.952**	0.949	0.033	0.899	0.903
FCL-COD	B	**0.023**	**0.945**	**0.960**	**0.043**	**0.872**	**0.835**	**0.020**	0.942	**0.954**	0.030	0.907	0.911

Figure 2. Qualitative Comparison Showing More Cases of FCL-COD vs. Competing Methods

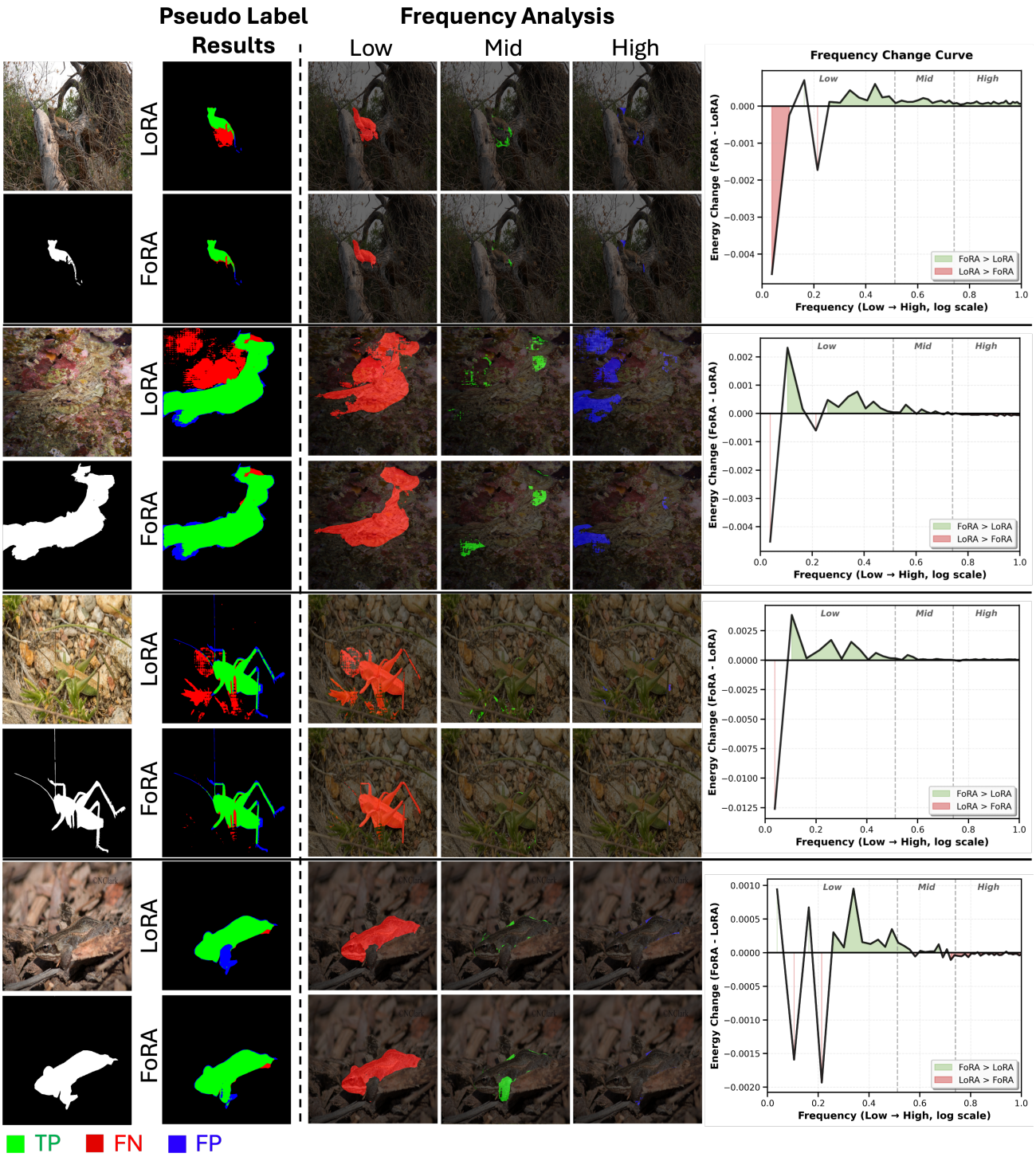

Figure 3. Visual and Spectral Analysis of FoRA (Part 1). The left columns compare the pseudo-labels generated by LoRA and FoRA, showing FoRA's superior structural integrity. The right columns visualize feature responses in frequency domains and the energy change curves, demonstrating FoRA's ability to selectively enhance boundary-related high-frequency components (green areas) while adjusting low-frequency responses.

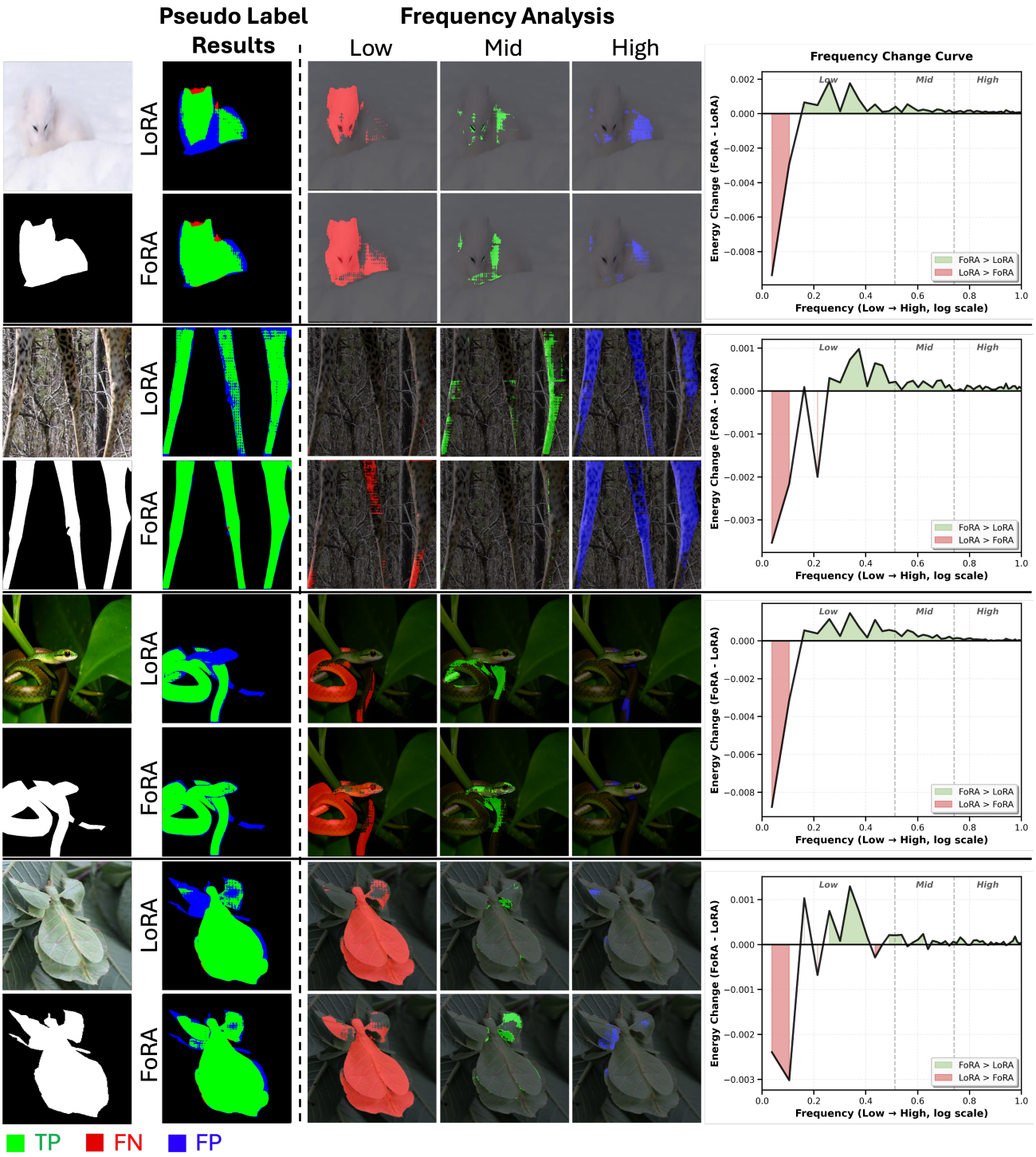

Figure 4. Visual and Spectral Analysis of FoRA (Part 2). Additional examples comparing LoRA and FoRA pseudo-labels, further validating FoRA's effectiveness in suppressing background noise and preserving object structure across diverse camouflaged scenarios.

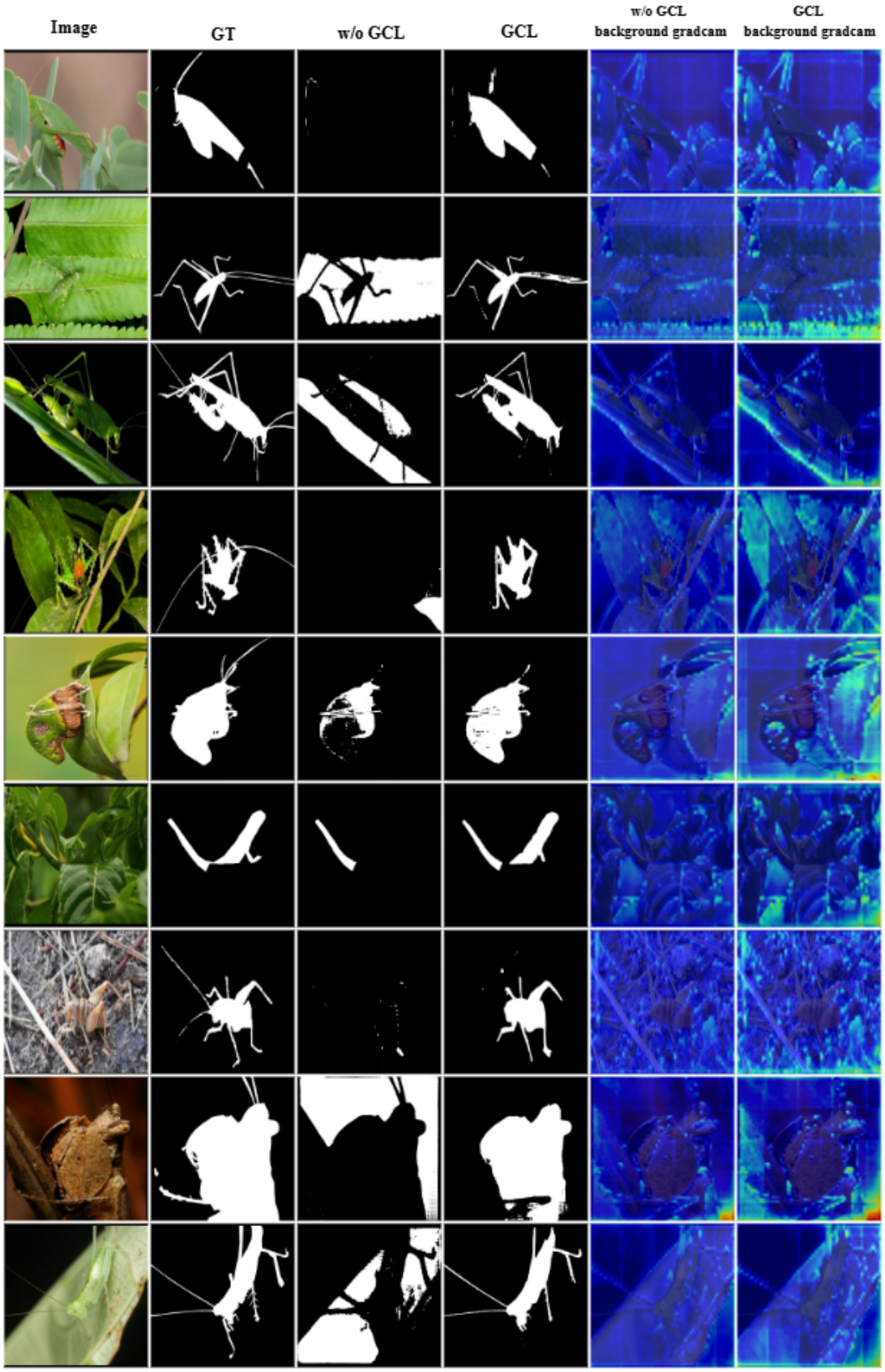

Figure 5. Grad-CAM Visual Comparison with and without GCL Module across Multiple Camouflaged Samples

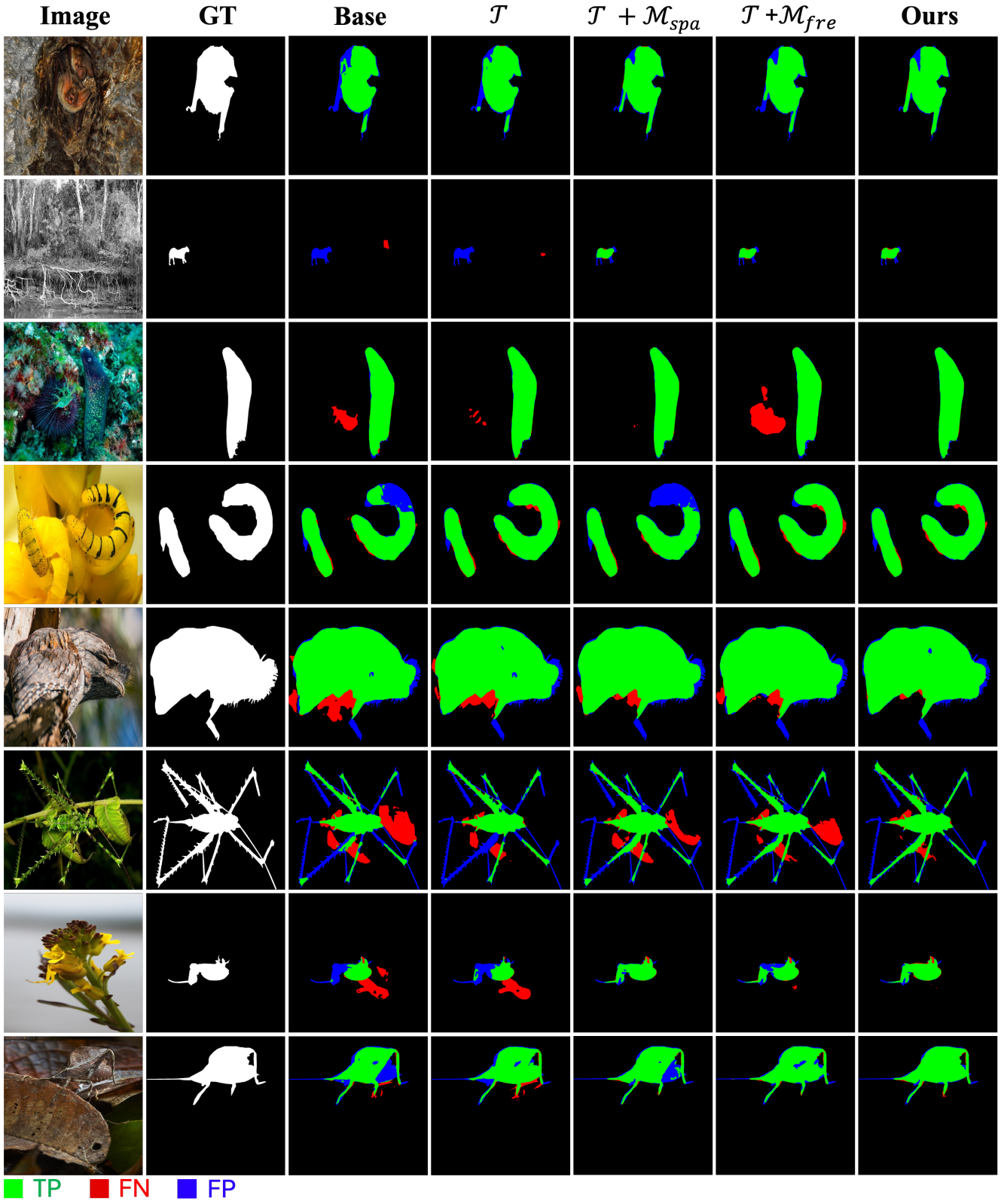

Figure 6. Visual Ablation Study of the MSFA Module. From left to right: Image, Ground Truth (GT), Baseline, Teacher Supervision ($\mathcal{T}$), $\mathcal{T}$ with Spatial Attention ($\mathcal{T} + \mathcal{M}_{spa}$), $\mathcal{T}$ with Frequency Attention ($\mathcal{T} + \mathcal{M}_{fre}$), and our full model. Green, red, and blue pixels represent True Positives (TP), False Negatives (FN), and False Positives (FP), respectively. The progressive reduction in error regions (red/blue) and increase in correct predictions (green) validates the complementary benefits of spatial and frequency attention.

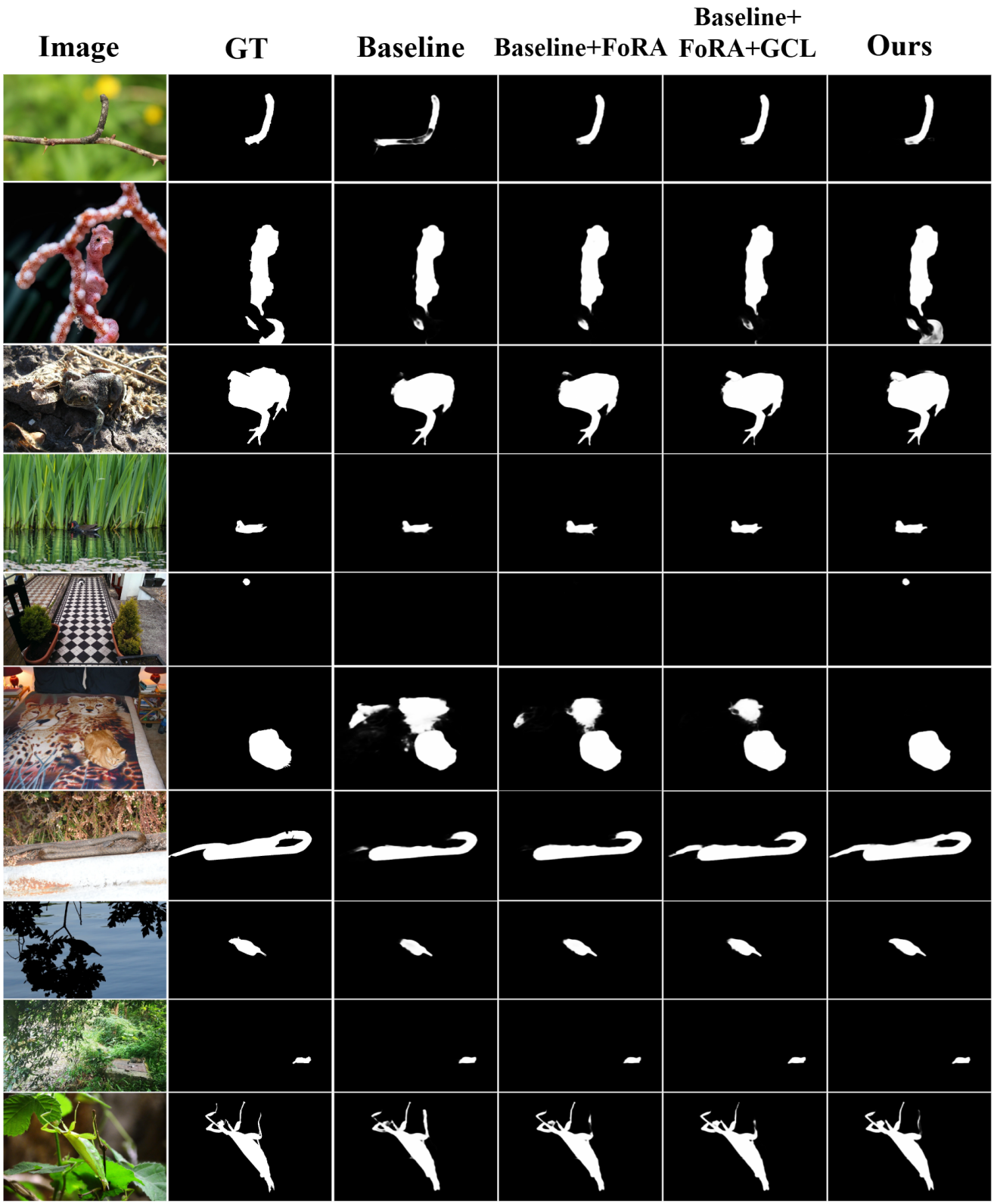

Figure 7. Qualitative Results of Progressive Module Integration in FCL-COD

References

[1] H. Chen, P. Wei, G. Guo, et al. Sam-cod+: Sam-guided unified framework for weakly-supervised camouflaged object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024. 3

[2] T. Chen, L. Zhu, C. Ding, et al. Sam fails to segment anything?–sam-adapter: Adapting sam in underperformed scenes: Camouflage, shadow, and more. *arXiv preprint arXiv:2304.09148*, 2023. 3

[3] D.-P. Fan, G.-P. Ji, G. Sun, et al. Camouflaged object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2777–2787, 2020. 3

[4] C. He, K. Li, Y. Zhang, et al. Camouflaged object detection with feature decomposition and edge reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22046–22055, 2023. 3

[5] C. He, K. Li, Y. Zhang, et al. Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. *Advances in Neural Information Processing Systems*, 36, 2024. 3

[6] R. He, Q. Dong, J. Lin, et al. Weakly-supervised camouflaged object detection with scribble annotations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 781–789, 2023. 3

[7] X. Hu, S. Wang, X. Qin, et al. High-resolution iterative feedback network for camouflaged object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 881–889, 2023. 3

[8] Z. Huang, H. Dai, T.-Z. Xiang, et al. Feature shrinkage pyramid for camouflaged object detection with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5557–5566, 2023. 3

[9] G.-P. Ji, D.-P. Fan, Y.-C. Chou, et al. Deep gradient learning for efficient camouflaged object detection. *Machine Intelligence Research*, 20(1):92–108, 2023. 3

[10] Q. Jia, S. Yao, Y. Liu, et al. Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4713–4722, 2022. 3

[11] A. Kirillov, E. Mintun, N. Ravi, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 3

[12] A. Li, J. Zhang, Y. Lv, et al. Uncertainty-aware joint salient object and camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10071–10081, 2021. 3

[13] Y. Liu, H. Li, J. Cheng, et al. Mscaf-net: A general framework for camouflaged object detection via learning multi-scale context-aware features. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(9):4934–4947, 2023. 3

[14] Y. Lv, J. Zhang, Y. Dai, et al. Simultaneously localize, segment and rank the camouflaged objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11591–11601, 2021. 3

[15] H. Mei, G.-P. Ji, Z. Wei, et al. Camouflaged object segmentation with distraction mining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8772–8781, 2021. 3

[16] Y. Pang, X. Zhao, T.-Z. Xiang, et al. Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 2160–2170, 2022. 3

[17] Y. Sun, G. Chen, T. Zhou, et al. Context-aware cross-level fusion network for camouflaged object detection. *arXiv preprint arXiv:2105.12555*, 2021. 3

[18] Y. Sun, S. Wang, C. Chen, et al. Boundary-guided camouflaged object detection. *arXiv preprint arXiv:2207.00794*, 2022. 3

[19] H. Xing, S. Gao, Y. Wang, et al. Go closer to see better: Camouflaged object detection via object area amplification and figure-ground conversion. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(10):5444–5457, 2023. 3

[20] F. Yang, Q. Zhai, X. Li, et al. Uncertainty-guided transformer reasoning for camouflaged object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4146–4155, 2021. 3

[21] B. Yin, X. Zhang, D.-P. Fan, et al. Camoformer: Masked separable attention for camouflaged object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 3

[22] S. Yu, B. Zhang, J. Xiao, et al. Structure-consistent weakly supervised salient object detection with local saliency coherence. In *Proceedings of the AAAI conference on artificial intelligence*, pages 3234–3242, 2021. 3

[23] Q. Zhai, X. Li, F. Yang, et al. Mutual graph learning for camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12997–13007, 2021. 3

[24] H. Zhu, P. Li, H. Xie, et al. I can find you! boundary-guided separated attention network for camouflaged object detection. In *Proceedings of the AAAI conference on artificial intelligence*, pages 3608–3616, 2022. 3