HFS-SAM2: Segment Anything Model 2 With High-Frequency Feature Supplementation for Camouflaged Object Detection

Zihuang Wu, Xinyu Xiong, *Member, IEEE*, Guangwei Gao, *Senior Member, IEEE*, Hongwei Li, and Hua Chen

***Abstract*—Camouflaged Object Detection (COD) aims to identify objects seamlessly blended with their backgrounds. While effective solutions exist for camouflaged animals, detecting camouflaged plants presents unique challenges and remains an open problem. This work introduces a novel Plant Camouflage Detection (PCD) method leveraging the Segment Anything Model 2 (SAM2). Our approach enhances the vanilla SAM2 decoder with specialized frequency-aware modules to improve the performance on PCD. Specifically, we employ a laplacian pyramid to extract high-frequency image components and introduce a High-Frequency Supplementation (HFS) module to enhance crucial spatial details for identifying camouflaged plants. The Multi-Scale Extraction (MSE) module is leveraged to capture rich multi-scale information, after which the features from the last three encoder layers are fused through a Cross-Layer Aggregation (CLA) module to obtain the aggregated high-level semantic features. A Semantic Gap Reduction (SGR) module is further proposed to bridge the semantic gap between high-level and shallow features during fusion. Finally, a Reverse Feature Mining (RFM) module is designed to highlight complementary regions and fine details. Extensive experiments on five datasets, encompassing both plant and animal camouflage detection, demonstrate the superior performance of our method compared to state-of-the-art approaches.**

***Index Terms*—Camouflaged object detection, segment anything model, high-frequency feature, deep learning.**

I. Introduction

CAMOUFLAGED Object Detection (COD) [1], [2], [3] has emerged as a critical area of research in computer vision, addressing the challenge of identifying objects that blend seamlessly into their surroundings. While the majority of COD studies focus on naturally camouflaged animals and artificially camouflaged objects like military gear or body art, the camouflage strategies of plants remain underexplored [4]. Much like animals, plants have evolved remarkable techniques to conceal themselves from predators, primarily herbivores, through strategies [5] such as background matching, disruptive coloration, masquerade, and decoration. These methods enable plants to survive in resource-scarce environments by mimicking elements of their surroundings, such as stones or twigs, or accumulating environmental materials like sand to reduce visibility. Such adaptations highlight the complexity and diversity of plant camouflage, distinct from the strategies employed by animals or synthetic objects.

Many effective methods have been developed in the field of universal camouflaged object detection. For instance, some methods draw inspiration from human or predator behavior [6], [7], [8], progressively refining results to improve detection accuracy. ZoomNet [8] adopts a zoom strategy to learn discriminative mixed-scale semantics through a scale integration unit and a hierarchical mixed-scale unit. Other approaches employ multi-task learning strategies [9], [10], [11] to enhance feature complementarity. For example, BSA-Net [9] introduces a boundary guider module that predicts object boundaries, improving boundary understanding and segmentation accuracy. Additionally, some research focuses on designing more effective attention mechanisms [12], [13] to enhance feature representation. PFNet [12] incorporates a distraction mining strategy to identify and suppress distraction regions, thereby improving detection robustness.

Despite the progress made by existing methods, there remains significant room for improvement in plant camouflage detection. To achieve better performance, this letter focuses on two key aspects. First, inspired by the emergence of vision foundation models [14], [15], [16], we explore adapting the Segment Anything Model 2 (SAM2) [17] for downstream segmentation tasks [18], [19], [20]. Unlike existing approaches that primarily focus on developing parameter-efficient fine-tuning strategies [21], [22], our approach centers on reconstructing the transformer decoder of SAM2, which is originally designed for video segmentation, to better align with the specific requirements of plant camouflage detection. Additionally, we propose to incorporate information beyond the RGB domain to more effectively distinguish foreground plants from the background. The high-frequency features extracted using

Received 25 March 2025; revised 10 July 2025; accepted 11 July 2025. Date of publication 16 July 2025; date of current version 29 July 2025. The associate editor coordinating the review of this article and approving it for publication was Prof. Yongjie Li. *(Corresponding author: Hua Chen.)*

Zihuang Wu, Hongwei Li, and Hua Chen are with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022, China (e-mail: 1874wzh@gmail.com; lihongwei@jxnu.edu.cn; chenhua5752@hotmail.com).

Xinyu Xiong is with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510000, China, and also with the EZVIZ, Hikvision, Hangzhou 310000, China (e-mail: xiongxyowo@gmail.com).

Guangwei Gao is with the Intelligent Visual Information Perception Laboratory, Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210046, China (e-mail: csggao@gmail.com).

Our code is available at https://github.com/WZH0120/HFS-SAM2.

Digital Object Identifier 10.1109/LSP.2025.3589944

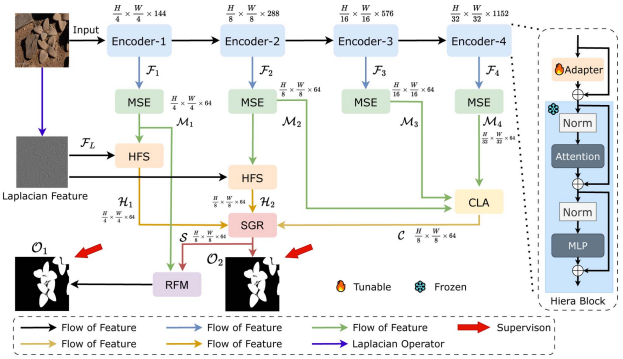

Fig. 1. The architecture of our proposed HFS-SAM2. The Adapter is composed of a "MLP-GeLU-MLP" structure.

Laplacian operations capture valuable boundary-related details, which help to more accurately identify plant edges during the decoding stage. In summary, our main contributions are as follows:

- To the best of our knowledge, we are the first to integrate the latest Segment Anything Model 2 (SAM2) with a frequency-aware decoder.
- By introducing the High-Frequency Supplementation (HFS) module, Multi-Scale Extraction (MSE) module, Cross-Layer Aggregation (CLA) module, Semantic Gap Reduction (SGR) module, and Reverse Feature Mining (RFM) module, our HFS-SAM2 effectively learns high-frequency features within multi-scale contexts while leveraging powerful representations from the foundation segmentation model to achieve accurate segmentation.
- Extensive experiments on five publicly available datasets, PlantCAMO, CAMO, CHAMELEON, COD10 K, and NC4K, demonstrate the state-of-the-art performance of our approach.

II. Method

A. Overview

Fig. 1 illustrates the architecture of HFS-SAM2, the decoder of which consists of five main modules: Multi-scale Extraction (MSE) module, Cross-Layer Aggregation (CLA) module, High-frequency Supplementation (HFS) module, Semantic Gap Reduction (SGR) module, and Reverse Feature Mining (RFM) module. For the SAM2 Hiera [23] encoder, we first freeze all its original parameters. Then, we insert an Adapter [24] module before each Hiera block of the SAM2 encoder, and these newly inserted Adapter modules are learnable. Since the Adapter is a simple MLP bottleneck structure with a significantly smaller number of parameters compared to the original encoder, this design allows us to fine-tune SAM2 efficiently. Given an image $I \in \mathbb{R}^{H\times W\times 3}$, the image is processed by the adapter-enhanced encoder to generate four hierarchical features $\mathcal{F}_i \in \mathbb{R}^{\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C_i}$, where $i \in \{1,2,3,4\}$ and $C_i \in \{144, 288, 576, 1152\}$. We apply the Laplacian operator to the image to obtain the Laplacian features $\mathcal{F}_L$. Through the collaborative operation of these five modules, we obtain the prediction results $\mathcal{O}_1$ and $\mathcal{O}_2$, with $\mathcal{O}_1$ being the final output. Next, we'll describe the proposed modules in detail.

TABLE I
Statistical Results of Various Methods on PlantCamo Dataset. The Best Results are Highlighted in **bold**

Method	Trainable Params	PlantCamo (250)			
		$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$E_\varphi^{ad} \uparrow$	$M \downarrow$
ZoomNet [8]	32.4M	0.798	0.680	0.874	0.049
SINetv2 [6]	24.9M	0.801	0.678	0.873	0.050
HitNet [30]	25.7M	0.854	0.794	0.929	0.034
GLCONet [31]	34.0M	0.857	0.776	0.909	0.036
PCNet [4]	25.7M	0.880	0.818	**0.937**	0.028
Ours	4.9M	**0.897**	**0.840**	0.934	**0.025**

B. Multi-Scale Extraction Module

Camouflaged object detection involves high diversity in scale and morphology, which brings significant challenges. To address this, we propose the Multi-Scale Extraction (MSE) module, specifically designed to enhance the model's ability to perceive targets of various sizes and shapes. The MSE module employs a multi-branch structure with dilated convolutions of different receptive fields, enabling effective capture of rich multi-scale information. This design not only improves adaptability to diverse targets but also reduces computational complexity through efficient feature compression. As depicted in Fig. 2, MSE consists of multiple branches with varying receptive field sizes, designed to capture multi-scale information:

$$\begin{aligned}
\mathcal{F}_i^1 &= \mathrm{Conv}_{1\times1}(\mathcal{F}_i),\\
\mathcal{F}_i^2 &= \mathrm{DConv}_{3,3}(\mathrm{Conv}_{3\times1}(\mathrm{Conv}_{1\times3}(\mathrm{Conv}_{1\times1}(\mathcal{F}_i)))),\\
\mathcal{F}_i^3 &= \mathrm{DConv}_{3,5}(\mathrm{Conv}_{5\times1}(\mathrm{Conv}_{1\times5}(\mathrm{Conv}_{1\times1}(\mathcal{F}_i)))),\\
\mathcal{F}_i^4 &= \mathrm{DConv}_{3,7}(\mathrm{Conv}_{7\times1}(\mathrm{Conv}_{1\times7}(\mathrm{Conv}_{1\times1}(\mathcal{F}_i)))),\\
\mathcal{F}_i^5 &= \mathrm{Conv}_{1\times1}(\mathrm{MaxPool}_{3,1}(\mathcal{F}_i)),\\
\mathcal{M}_i &= \mathrm{Conv}_{3\times3}(\mathrm{Cat}[\mathcal{F}_i^1,\mathcal{F}_i^2,\mathcal{F}_i^3,\mathcal{F}_i^4,\mathcal{F}_i^5]) \oplus \mathcal{F}_i^1, \qquad (1)
\end{aligned}$$

where $\mathcal{F}_i$ denotes the encoder feature, $\mathcal{F}_i^j$ represents different branch in the MSE module. $\mathrm{Conv}_{1\times1}(\cdot)$ represents a 1×1 convolution, $\mathrm{DConv}_{3,5}(\cdot)$ denotes a 3×3 dilated convolution with a dilation rate of 5, $\mathrm{MaxPool}_{3,1}(\cdot)$ refers to max pooling with a kernel size of 3 and a stride of 1, and $\mathrm{Cat}[\cdot]$ denotes the concatenation operation along channel dimension.

C. Cross-Layer Aggregation Module

Deep features contain rich semantic information, which is essential for accurate camouflaged object detection. To obtain more comprehensive and robust semantic representations, we propose a Cross-Layer Aggregation (CLA) module that aggregates features from multiple high-level encoder layers. By fusing deep semantic features, CLA enhances the model's ability to localize and identify camouflaged objects in complex backgrounds. As shown in Fig. 2, CLA can be represented as:

$$\begin{aligned}
\mathcal{M}_4' &= \mathfrak{f}_3(\mathrm{Up}(\mathcal{M}_4)),\\
\mathcal{M}_3' &= \mathrm{Cat}[\mathfrak{f}_3(\mathcal{M}_3 \otimes \mathcal{M}_4' \oplus \mathcal{M}_3), \mathcal{M}_4'],
\end{aligned}$$

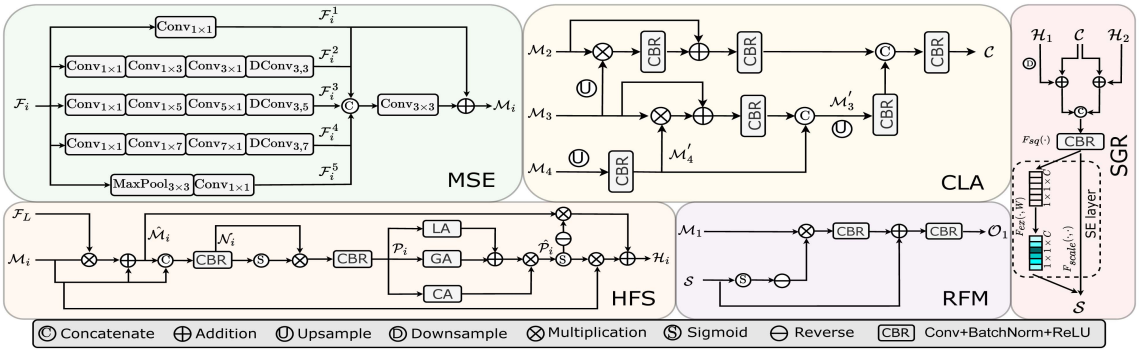

Fig. 2. The architecture of the proposed different modules, including Multi-Scale Extraction (MSE) module, Cross-Layer Aggregation (CLA) module, High-Frequency Supplementation (HFS) module, Semantic Gap Reduction (SGR) module, and Reverse Feature Mining (RFM) module.

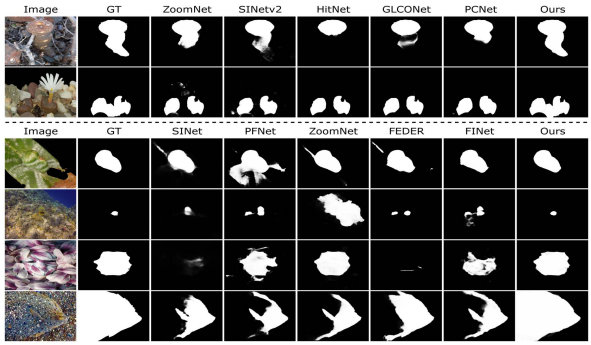

Fig. 3. Qualitative comparison with SOTA methods on PlantCamo Dataset (up) and general COD datasets (down).

$$\mathcal{C} = \mathfrak{f}_3(\mathrm{Cat}[\mathfrak{f}_3(\mathfrak{f}_3(\mathcal{M}_2 \otimes \mathrm{Up}(\mathcal{M}_3)) \oplus \mathcal{M}_2), \mathfrak{f}_3(\mathrm{Up}(\mathcal{M}_3'))]), \tag{2}$$

where $\mathcal{M}_4$ is obtained from $\mathcal{F}_4$ after being processed by the MSE module, $\mathfrak{f}_3(\cdot)$ represents a 3×3 convolution + BatchNorm + ReLU, and $\mathrm{Up}(\cdot)$ denotes $2\times$ upsampling via bilinear interpolation.

D. High-Frequency Supplementation Module

The multi-head attention mechanism in the Transformer encoder tends to capture low-frequency information, while high-frequency features are crucial in camouflage object detection as they help the model better distinguish the foreground from the background. Inspired by this, we first use the Laplace operator to extract the high-frequency components from the original image. Then, we propose a High-Frequency Supplementation (HFS) module to mine high-frequency components and supplement the missing high-frequency details in the encoder, thereby enhancing the model's information. As shown in Fig. 2, HFS can be expressed as:

$$\begin{aligned}
\hat{\mathcal{M}}_i &= \mathcal{M}_i \otimes \mathcal{F}_L \oplus \mathcal{M}_i, \\
\mathcal{N}_i &= \mathfrak{f}_3(Cat[\hat{\mathcal{M}}_i, \mathcal{M}_i]), \\
\mathcal{P}_i &= \mathfrak{f}_3(\sigma(\mathcal{N}_i) \otimes \mathcal{N}_i), \\
\hat{\mathcal{P}}_i &= \sigma((\mathrm{LA}(\mathcal{P}_i) \oplus \mathrm{GA}(\mathcal{P}_i)) \otimes \mathrm{CA}(\mathcal{P}_i)), \\
\mathcal{H}_i &= \hat{\mathcal{P}}_i \otimes \mathcal{M}_i \oplus (\ominus(\hat{\mathcal{P}}_i) \otimes \hat{\mathcal{M}}_i),
\end{aligned} \tag{3}$$

where $\sigma(\cdot)$ denotes the Sigmoid activation function, $\mathrm{LA}(\cdot)$ represents local attention, $\mathrm{GA}(\cdot)$ denotes global attention, $\mathrm{CA}(\cdot)$ refers to channel attention, and $\ominus(\cdot)$ represents a reverse operation.

E. Semantic Gap Reduction Module

After obtaining high-level semantic features through CLA and high-frequency component-enhanced shallow features through HFS, a significant semantic gap exists between these two types of features. Simply concatenating or adding them may lead to performance degradation of the model. Therefore, we propose a Semantic Gap Reduction (SGR) module to adaptively fuse these features, enabling effective feature complementation. As shown in Fig. 2, SGR can be expressed as:

$$\mathcal{S} = \mathrm{SE}(\mathfrak{f}_1(\mathrm{Cat}[\mathcal{C} \oplus \mathrm{Down}(\mathcal{H}_1), \mathcal{C} \oplus \mathcal{H}_2])), \tag{4}$$

where $\mathrm{SE}(\cdot)$ represents the Squeeze-and-Excitation module [25], $\mathfrak{f}_1(\cdot)$ represents a 1×1 convolution + BatchNorm + ReLU, and $\mathrm{Down}(\cdot)$ denotes downsampling achieved by max pooling with a kernel size of 3 and a stride of 2.

F. Reverse Feature Mining Module

To enhance the model's ability to handle boundary regions, we employ a reverse attention mechanism to encourage the model to focus more on the background areas, mining potential complementary details from the background. As shown in Fig. 2, RFM can be expressed as:

$$\mathcal{O}_1 = \mathfrak{f}_1(\mathfrak{f}_1(\ominus(\sigma(\mathcal{S})) \otimes \mathcal{M}_1) \oplus \mathcal{S}). \tag{5}$$

where $\mathfrak{f}_1(\cdot)$ represents a 1×1 convolution + BatchNorm + ReLU.

G. Loss Function

Following [1], [26], we adopt a combination of weighted cross-entropy loss (L^{ω}_{BCE}) and weighted IoU loss (L^{ω}_{IoU}) as the loss function, the detailed implementation of which can be found in [26]. Specifically, the L^{ω}_{BCE} is used to measure the pixel-wise classification error, where higher weights are assigned to pixels that are more difficult to classify. The L^{ω}_{IoU} evaluates the overlap between the predicted segmentation and the ground truth, with additional emphasis on challenging pixels to further improve

TABLE II
STATISTICAL RESULTS WITH VARIOUS METHODS ON GENERAL COD DATASETS, WHERE THE TEST SETS OF CAMO, CHAMELEON, COD10K, AND NC4K CONTAIN 250, 76, 2026, AND 4121 IMAGES RESPECTIVELY. THE BEST RESULTS ARE HIGHLIGHTED IN **bold**

Method	Trainable Params	CAMO (250)				CHAMELEON (76)				COD10K (2,026)				NC4K (4,121)			
		$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$E_\varphi^{ad} \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$E_\varphi^{ad} \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$E_\varphi^{ad} \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$E_\varphi^{ad} \uparrow$	$M \downarrow$
SINet [1]	48.9M	0.751	0.606	0.834	0.100	0.869	0.740	0.899	0.044	0.771	0.551	0.797	0.051	0.808	0.723	0.883	0.058
PFNet [12]	46.5M	0.782	0.695	0.855	0.085	0.882	0.810	0.942	0.033	0.800	0.660	0.868	0.040	0.829	0.745	0.894	0.053
ZoomNet [8]	32.4M	0.820	0.752	0.883	0.066	0.902	0.845	**0.952**	0.023	0.838	0.729	0.893	0.029	0.853	0.784	0.907	0.043
FEDER [13]	42.1M	0.802	0.738	0.877	0.071	0.887	0.834	0.943	0.030	0.822	0.716	0.901	0.032	0.847	0.789	0.913	0.044
SAM [32]	-	0.684	0.606	0.689	0.132	0.727	0.639	0.736	0.081	0.783	0.701	0.800	0.049	0.767	0.696	0.778	0.078
DSAM [21]	-	0.832	0.794	-	0.061	-	-	-	-	0.846	0.760	-	0.033	0.871	0.826	-	0.040
FINet [33]	3.7M	0.828	0.752	0.890	0.065	0.883	0.808	0.929	0.031	0.817	0.686	0.883	0.034	0.847	0.771	0.904	0.047
PCNet [4]	25.7M	0.849	0.792	0.908	0.056	0.896	0.839	0.939	0.026	0.843	0.742	0.908	0.029	0.871	0.816	0.925	0.038
VSCode [34]	60.6M	0.873	0.820	**0.928**	0.046	-	-	-	-	0.869	0.780	**0.929**	**0.023**	0.891	0.841	**0.939**	0.032
GLCONet [31]	34.0M	0.816	0.748	0.882	0.069	-	-	-	-	0.825	0.714	0.891	0.033	0.853	0.791	0.914	0.043
Ours	4.9M	**0.883**	**0.839**	0.922	**0.045**	**0.909**	**0.849**	0.943	**0.023**	**0.879**	**0.789**	0.913	**0.023**	**0.900**	**0.849**	0.925	**0.030**

- denotes the results are unavailable in the original paper.

segmentation accuracy. We supervise the model's predictions $\mathcal{O}_1$ and $\mathcal{O}_2$, with $\mathcal{O}_1$ being the final output of the model. The loss function can be expressed as:

$$\mathcal{L} = \mathcal{L}_{BCE}^{\omega} + \mathcal{L}_{IoU}^{\omega},$$
$$\mathcal{L}_{total} = \sum_{i=1}^{i=2} \mathcal{L}(GT, \mathcal{O}_i). \tag{6}$$

III. EXPERIMENTS

A. Experimental Settings

1) Datasets: Consistent with the dataset setup in [4], we evaluate the performance of our model and comparative models on the PlantCamo dataset [4] for plant camouflage segmentation. Specifically, the PlantCamo dataset contains a total of 1250 images, with 1000 images used for training and the remaining 250 images for testing. Furthermore, we also conduct experiments on four widely used general COD datasets, including CHAMELEON [27], COD10K [1], NC4K [28], and CAMO [29], where 1,000 images from CAMO and 3,040 images COD10 K make up the training set.

2) Evaluation Metrics: Following [34], we adopt four metrics to evaluate performance, including Structure-measure (S_α) [35], weighted F-measure (F_β^ω) [36], adaptive E-measure (E_φ^{ad}) [37], and mean absolute error (M).

3) Implementation Details: We implement our model using PyTorch and conduct experiments on a 24 GB NVIDIA RTX 3090. We use the SAM2-Hiera-L as the encoder. Following the setup in [4], we scale the input images during training to 704× 704, use the AdamW optimizer, and set the number of epochs to 150. Additionally, the learning rate is set to 1e-3 and the weight decay is set to 5e-4. For data augmentation, similar to [4], we apply random rotation, random horizontal flipping, and vertical flipping.

B. Quantitative Comparison

As shown in Table I, our HFS-SAM2 model achieves the best performance on the plant COD dataset. Specifically, compared to the second-best PCNet, HFS-SAM2 demonstrates improvements of 1.7%, 2.2%, and 0.003 in Structure-measure (S_α), weighted F-measure (F_β^ω), and mean absolute error (M), respectively. Furthermore, to further evaluate the learning capacity and generalization ability of HFS-SAM2, we conducted additional experiments on four commonly used COD datasets. As presented in Table II, our model achieves the highest S_α scores on CAMO, CHAMELEON, COD10K, and NC4K, with values of 0.883, 0.909, 0.879, and 0.900, respectively. Moreover, HFS-SAM2 shows significant improvements over the original SAM and the SAM-based method DSAM. These results clearly highlight the superiority of our HFS-SAM2.

TABLE III
ABLATION STUDY OF DIFFERENT MODULES

Method	PlantCamo (250)			
	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$E_\varphi^a d \uparrow$	$M \downarrow$
w/o MSE	0.880	0.810	0.916	0.029
w/o CLA	0.890	0.828	0.925	0.027
w/o HFS	0.890	0.825	0.928	0.026
w/o SGR	0.890	0.822	0.916	0.027
w/o RFM	0.891	0.829	0.928	**0.025**
Ours	**0.897**	**0.840**	**0.934**	**0.025**

C. Qualitative Comparison

Fig. 3 present the qualitative comparison results on the PlantCamo and the commonly used COD datasets, respectively. In these challenging scenarios, our model achieves superior performance, while other methods fail to provide complete predictions (e.g., row 1 and 2 of Fig. 3 (up)), or are confounded by the surrounding background (e.g., row 1 and 2 of Fig. 3 (down)).

D. Inference Speed Analysis

Our method achieves an inference speed of 16 FPS, which is more efficient than another SAM-based method, DSAM [21], with a speed of 3.8 FPS.

E. Ablation Study

As shown in Table III, we validate the importance of MSE, CLA, HFS, SGR, and RFM by systematically removing each component. Specifically, when HFS was removed, S_α on PlantCamo decreased by 0.7%, indicating that HFS effectively enhances the model's performance by adaptively utilizing high-frequency components to supplement detail information. Moreover, it is clear that the removal of any module leads to a decline in the model's performance.

IV. CONCLUSION

In this letter, we propose a frequency-aware network, HFS-SAM2, based on SAM2 for plant camouflage detection. Extensive experiments on five datasets demonstrate that HFS-SAM2 exhibits strong competitiveness compared to state-of-the-art methods.

REFERENCES

[1] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2777–2787.

[2] C. Li, G. Jiao, Y. Wu, and W. Zhao, "Camouflaged instance segmentation from global capture to local refinement," *IEEE Signal Process. Lett.*, vol. 31, pp. 661–665, 2024.

[3] J. Wu, W. Liang, F. Hao, and J. Xu, "Mask-and-edge co-guided separable network for camouflaged object detection," *IEEE Signal Process. Lett.*, vol. 30, pp. 748–752, 2023.

[4] J. Yang, Q. Wang, F. Zheng, P. Chen, A. Leonardis, and D.-P. Fan, "Plantcamo: Plant camouflage detection," 2024, *arXiv:2410.17598*.

[5] Y. Niu, H. Sun, and M. Stevens, "Plant camouflage: Ecology, evolution, and implications," *Trends Ecol. Evol.*, vol. 33, no. 8, pp. 608–618, 2018.

[6] D.-P. Fan, G.-P. Ji, M.-M. Cheng, and L. Shao, "Concealed object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6024–6042, Oct. 2022.

[7] H. Xing, S. Gao, Y. Wang, X. Wei, H. Tang, and W. Zhang, "Go closer to see better: Camouflaged object detection via object area amplification and figure-ground conversion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 10, pp. 5444–5457, Oct. 2023.

[8] Y. Pang, X. Zhao, T.-Z. Xiang, L. Zhang, and H. Lu, "Zoom in and out: A mixed-scale triplet network for camouflaged object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 2160–2170.

[9] H. Zhu et al., "I can find you! Boundary-guided separated attention network for camouflaged object detection," in *Proc. Conf. AAAI Artif. Intell.*, 2022, pp. 3608–3616.

[10] Q. Zhang, X. Sun, Y. Chen, Y. Ge, and H. Bi, "Attention-induced semantic and boundary interaction network for camouflaged object detection," *Comput. Vis. Image Understanding*, vol. 233, 2023, Art. no. 103719.

[11] Y. Sun, S. Wang, C. Chen, and T.-Z. Xiang, "Boundary-guided camouflaged object detection," in *Proc. Int. Joint Conferences Artif. Intell.*, 2022, pp. 1335–1341.

[12] H. Mei, G.-P. Ji, Z. Wei, X. Yang, X. Wei, and D.-P. Fan, "Camouflaged object segmentation with distraction mining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8772–8781.

[13] C. He et al., "Camouflaged object detection with feature decomposition and edge reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 22046–22055.

[14] A. Kirillov et al., "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 4015–4026.

[15] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Commun.*, vol. 15, no. 1, 2024, Art. no. 654.

[16] W. Li, X. Xiong, P. Xia, L. Ju, and Z. Ge, "TP-DRSeg: Improving diabetic retinopathy Lesion segmentation with explicit text-prompts assisted SAM," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2024, pp. 743–753.

[17] N. Ravi et al., "Sam 2: Segment anything in images and videos," 2024, *arXiv:2408.00714*.

[18] J. Zhu, A. Hamdi, Y. Qi, Y. Jin, and J. Wu, "Medical sam 2: Segment medical images as video via segment anything model 2," 2024, *arXiv:2408.00874*.

[19] X. Xiong et al., "Sam2-UNet: Segment anything 2 makes strong encoder for natural and medical image segmentation," 2024, *arXiv:2408.08870*.

[20] Y. Zhou, G. Sun, Y. Li, L. Benini, and E. Konukoglu, "When sam2 meets video camouflaged object segmentation: A comprehensive evaluation and adaptation," 2024, *arXiv:2409.18653*.

[21] Z. Yu, X. Zhang, L. Zhao, Y. Bin, and G. Xiao, "Exploring deeper! Segment anything model with depth perception for camouflaged object detection," in *Proc. ACM Int. Conf. Multimedia*, 2024, pp. 4322–4330.

[22] X. Xiong, C. Wang, W. Li, and G. Li, "Mammo-SAM: Adapting foundation segment anything model for automatic breast mass segmentation in whole mammograms," in *Proc. Int. Workshop. Mach. Learn. Med. Imag.*, 2023, pp. 176–185.

[23] C. Ryali et al., "Hiera: A hierarchical vision transformer without the bells-and-whistles," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 29441–29454.

[24] N. Houlsby et al., "Parameter-efficient transfer learning for nlp," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2790–2799.

[25] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[26] J. Wei, S. Wang, and Q. Huang, "F^3Net: Fusion, feedback and focus for salient object detection," in *Proc. Conf. AAAI Artif. Intell.*, 2020, pp. 12321–12328.

[27] P. Skurowski, H. Abdulameer, J. Błaszczyk, T. Depta, A. Kornacki, and P. Kozieł, "Animal camouflage analysis: Chameleon database," *Unpublished Manuscript*, vol. 2, no. 6, 2018, Art. no. 7.

[28] Y. Lv et al., "Simultaneously localize, segment and rank the camouflaged objects," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 11591–11601.

[29] T.-N. Le, T. V. Nguyen, Z. Nie, M.-T. Tran, and A. Sugimoto, "Anabranch network for camouflaged object segmentation," *Comput. Vis. Image Understanding*, vol. 184, pp. 45–56, 2019.

[30] X. Hu et al., "High-resolution iterative feedback network for camouflaged object detection," in *Proc. Conf. AAAI Artif. Intell.*, 2023, pp. 881–889.

[31] Y. Sun, H. Xuan, J. Yang, and L. Luo, "GLCONet: Learning multisource perception representation for camouflaged object detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 7, pp. 13262–13275, Jul. 2025.

[32] W. Liang, J. Wu, Y. Wu, X. Mu, and J. Xu, "FINet: Frequency injection network for lightweight camouflaged object detection," *IEEE Signal Process. Lett.*, vol. 31, pp. 526–530, 2024.

[33] Z. Luo et al., "Vscode: General visual salient and camouflaged object detection with 2D prompt learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 17169–17180.

[34] G. Chen, S.-J. Liu, Y.-J. Sun, G.-P. Ji, Y.-F. Wu, and T. Zhou, "Camouflaged object detection via context-aware cross-level fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6981–6993, Oct. 2022.

[35] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2017, pp. 4548–4557.

[36] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps?," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 248–255.

[37] D.-P. Fan, G.-P. Ji, X. Qin, and M.-M. Cheng, "Cognitive vision inspired object segmentation metric and loss function," *Scientia Sinica Informationis*, vol. 6, no. 6, 2021, Art. no. 5.