Few-Shot Instance Segmentation: An Exploration in the Frequency Domain for Camouflage Instances

Thanh-Danh Nguyen[1,2], Hung-Phu Cao[3], Thanh Duc Ngo[1,2], Vinh-Tiep Nguyen[†1,2], and Tam V. Nguyen[4]

[1]University of Information Technology, Ho Chi Minh City, Vietnam
[2]Vietnam National University, Ho Chi Minh City, Vietnam
[3]Endava Vietnam, Ho Chi Minh City, Vietnam
[4]University of Dayton, Dayton, OH 45469, United States
{*danhnt, thanhnd, tiepnv*}*@uit.edu.vn, caohungphu@hotmail.com, tamnguyen@udayton.edu*, [†]*corresponding author*

***Abstract*—Few-shot instance segmentation is an intense yet essential task, particularly in camouflaged scenarios where visual ambiguity between foreground and background makes instance-level recognition more difficult. Prior approaches primarily focused on image augmentations in the color space domain to provide diverse perspective information to the segmentation models. However, this type of augmentation often fails to capture the full range of visual characteristics needed for robust generalization, particularly in camouflage images, due to the limited similar representation in the color space domain. To this end, we tackle this gap by exploiting a novel approach to augment and enhance image features in the derivative frequency domain. Accordingly, we propose a novel framework tailored for few-shot camouflage instance segmentation via the instance-aware frequency-based augmentation, dubbed FS-CAMOFreq, to enhance image diversity while preserving semantic structure, thereby improving the ability of the few-shot segmentor to learn from limited data. Extensive experiments on the challenging CAMO-FS benchmark demonstrate that our approach achieves superior performance compared to state-of-the-art baselines. Code can be found at https://github.com/danhntd/FS-CAMOFreq.**

***Index Terms*—Few-shot Instance Segmentation, Camouflage Instance Segmentation, Frequency Domain Augmentation,**

I. Introduction

Image segmentation is a cornerstone task in computer vision, serving as a foundation for various applications ranging from autonomous driving and medical imaging to visual content analysis and robotics. Traditional segmentation approaches aim to delineate object regions from background scenes at the pixel level, offering either semantic understanding (semantic segmentation) or instance-level differentiation (instance segmentation). In practical settings, especially where real-time deployment or scalability is essential, utilizing large-scale pixel-wise annotations remains prohibitively expensive. Specifically in camouflage research, such applications in search and rescue jobs or military surveillance gained significant attention. This limitation has motivated the development of few-shot learning in segmentation, which focuses on learning segmentation tasks from a limited number of annotated samples [1]–[7]. Few-shot learning offers the potential to generalize across categories and domains with little supervision, making it particularly valuable for real-world applications, data-scarce scenarios, including camouflage contexts [4], [8].

Few-shot semantic segmentation (FSS) and few-shot instance segmentation (FSIS) share a fundamental challenge: learning to generalize from a limited number of annotated samples to a large and diverse set of unseen images. This constraint often leads to suboptimal performance, especially in instance-level tasks where accurate delineation of individual object boundaries is critical. Specifically in the case of camouflage instance segmentation (CIS) where the foreground instances blend in with the contextual background [4], [9]–[12]. In addition, the inherent complexity of FSIS, such as handling occlusions, fine-grained distinctions between instances, and intraclass variability, further exacerbates the data scarcity problem. To mitigate this, previous research has focused on data enrichment strategies to provide different perspectives of a sample to feed the understanding tasks. Such works on data augmentation operate primarily in the spatial or color domain, such as pixel-level transformations and photometric augmentations. However, these methods are often limited in their ability to capture the broader spectrum of visual variability as they only walk around in the color space. To this end, a novel approach is introduced to the field of image segmentation that utilizes the image adjustment in the frequency domain and offers a promising alternative, as it enables the manipulation of structural and textural components of images without compromising semantic integrity [13]–[15]. This spectral augmentation can introduce meaningful diversity into the training data, thereby improving generalization and robustness in few-shot instance segmentation tasks. Moreover, in camouflage research, the frequency-based augmentation is claimed to be potentially beneficial as it may enhance the camouflage features that look similar in the color space [13].

Prior work succeeded in this approach in a fully supervised manner [13], [16]–[18]; however, there is no work that evaluated its ability in few-shot learning to demonstrate the effectiveness under a scarce data context. Frequency-based augmentation has shown promise in improving generalization for visual tasks, but its potential remains underexplored in the context of FSIS for camouflage research. Motivated by this gap, we propose to incorporate frequency domain augmentation as a core component in the few-shot instance segmentation pipeline, aiming to enrich feature diversity and

mitigate overfitting in low-shot regimes. To summarize, our contributions in this work are threefold:

- Firstly, we propose a novel framework, dubbed FS-CAMOFreq, that exploits the frequency domain under an instance-aware enhancement to improve the few-shot camouflage instance segmentation.
- Secondly, we introduce a few-shot frequency-based image augmentation technique exploring the relative intra-class phase information in the frequency domain to adjust the visual appearance of the instances in the images, especially effective in the case of camouflage objects.
- Thirdly, we demonstrate the effectiveness of our proposed framework on the challenging CAMO-FS benchmark [4] via extensive experiments.

The remainder of this paper is organized as follows. Section II reviews related work on image segmentation in general and in the specific camouflage research. This section also reviews the few-shot camouflage instance segmentation task. Section III presents our proposed FS-CAMOFreq with details on each proposed component. In Section IV, we report our experiments and ablation studies to prove the effectiveness of our proposal. Finally, Section V concludes our work.

II. Related work

A. Image Segmentation

Semantic Segmentation. The traditional approach to address semantic segmentation mostly relied on convolutional neural network architectures, framing the task as a dense pixel-wise classification problem [19]–[21]. While CNNs proved effective at modeling local spatial features through hierarchical convolutional operations, their inherently limited receptive fields restricted their ability to capture long-range dependencies and holistic scene structure. To address this limitation, the emergence of transformer-based architectures marked a significant trend in semantic segmentation research. Models such as Segmenter [22], SegFormer [23], SeMask [24], and SARFormer [25] extend the Vision Transformer (ViT) framework [26] by incorporating the self-attention mechanism originally developed for natural language processing [27]. These architectures excel at modeling global context and relational dependencies across image regions, which are critical for accurate semantic interpretation in complex visual scenes. Despite these advancements, semantic segmentation inherently operates at the semantic pixel-wise level and does not differentiate between individual object instances. Consequently, it remains insufficient for applications that demand fine-grained instance-level understanding, such as human tracking and counting, visual reasoning, or scene decomposition.

Instance Segmentation. Instance segmentation is proposed to resolve the limitation of semantic segmentation when it is capable of differentiating individual instances in the same semantic classes. This task is often classified into two categories: one-stage and two-stage models. Such models as Mask R-CNN [28], Cascade R-CNN [29], PANet [30], HTC [31], and DetectoRS [32] are typically mentioned for the two-stage approach, where the accurate prediction ability dominates the real-time manner. However, recent research pays more attention to the one-stage approach, as the models are both efficient and convenient in end-to-end designs. Models such as YOLACT [33], SOLO [34], Mask2Former [35], FastInst [36], and OneFormer [37] unify detection and segmentation in a shared architecture. However, such aforementioned methods required abundant annotated data for training to achieve high accuracy, which is limited under the concept of camouflage animals [10], [38].

B. Few-shot Learning in Image Segmentation

To address the image segmentation under limited data conditions, the community has come up with the solution of few-shot learning, where the models only digest a modest amount of annotated data but still maintain accurate predictions on the segmentation mask.

Few-shot Semantic Segmentation. Early few-shot learning efforts primarily focused on semantic segmentation. Dong *et al.* [39] introduced a prototype learning component that extracts discriminative feature representations for class-level segmentation. Wang *et al.* [30] proposed a prototype alignment mechanism to learn class-specific prototypes for query segmentation. Liu *et al.* [40] advanced this line with a cross-reference network and a mask refinement module to enhance segmentation accuracy. Further, [41] presented context-aware prototype learning, while [42] explored generative modeling approaches. Transformer-based models have also emerged, including RefT by Han *et al.* [43] and DTN by Wang *et al.* [44], which enhance feature representation and segment objects directly from references. However, as aforementioned, such methods limit their capability in segmenting semantic objects without considering each individual instance.

C. Few-shot Camouflage Image Segmentation

Few-shot Instance Segmentation. In the instance-level setting, Meta R-CNN [40] extended Mask R-CNN [28] to jointly perform detection and segmentation. iMTFA [3] also relied on the common Mask R-CNN [28] to set up few-shot learning with the incremental approach. Nguyen *et al.* [2] proposed iFS-RCNN, introducing an incremental learning strategy for instance segmentation. Gao *et al.* [45] developed the DCFS framework, a decoupling classifier that effectively improves both detection and segmentation performance in few-shot scenarios. Han *et al.* propose RefT [43] as a simple and unified baseline for few-shot instance segmentation. MaskDiff [5] utilizes the architecture of a diffusion-based model to address the task. Such methods are proposed to serve the generic domain object detection or instance segmentation.

Few-shot Camouflage Image Segmentation. Recently, FS-CDIS of Nguyen *et al.* [4] stands out as the pioneer work proposing a few-shot learning framework for camouflage object detection and instance segmentation with contrastive components of instance triplet loss and instance memory storage. This framework is generalized on top of iFS-RCNN [2], iMTFA [3], and Mask R-CNN [28], thus presenting a

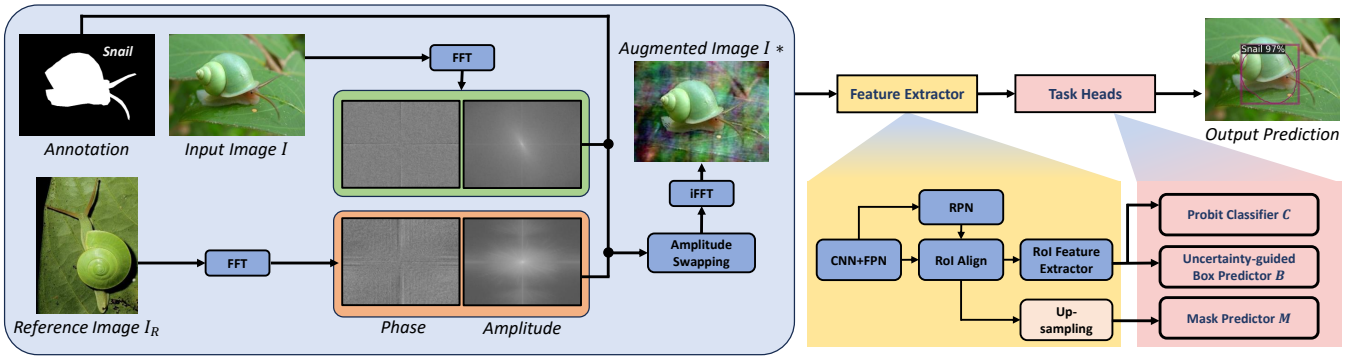

Fig. 1. Overview of our FS-CAMOFreq framework exploiting the instance-ware frequency-based enhancement in few-shot camouflage instance segmentation.

strong baseline for derivative search in this field. In this work, we inherit FS-CDIS [2], [4] to develop our frequency domain enhancement in few-shot camouflage instance segmentation.

D. Data Enhancement in Camouflage Image Segmentation

Traditional image data enhancement techniques rely on low-level feature augmentations, such as geometric transformations [46], [47] and color space transformations [48]–[50]. However, such methods may fall short in producing diverse samples that fully capture the spectrum of variations in the original dataset due to the limitation in visual representation. Indeed, an RGB-color image limits the appearance of the scene to three color channels. Thus, some information is not well presented, such as global structural patterns or scale of variations, which is better expressed under the frequency domain. Prior work exploited the effectiveness of this derivative domain in multiple visual tasks, including image classification [16], object detection [17], and image segmentation [18], [51]. Recently, the CamoFA [13] extended the idea to utilize frequency-based augmentation on camouflage research in a fully supervised approach, which is considered a drawback when it digests large amounts of annotated data. In this work, we leverage the frequency domain to transform the camouflage instances and resolve the context of limited camouflage samples.

III. Proposed Method

A. Few-Shot CIS Formulation.

In the task of few-shot learning, we have two phases of learning, which are the base phase and the fine-tuning phase. Corresponding to each phase, we have one set of base classes denoted as C_{base}, having a large amount of available annotated training data, and another disjoint set of novel classes denoted as C_{novel}, containing a small amount of training data. The small amount of data in C_{novel} is limited to a few samples. The target is to train the model to explore the knowledge in the base data C_{base} and predict well on novel data $C_{test} = C_{novel}$ [52], [53] or on both base and novel data $C_{test} = C_{base} \cup C_{novel}$ [3], [4], [54]. In few-shot classification, [53] introduces episodic training where the method sets up a series of episodes $E_i = (I_q, S_i)$ with S_i is a support set that contains N classes from $C_{train} = C_{novel} \cup C_{base}$ along with K examples per class (so-called N-way K-shot). A network is then trained to classify an input query image I_q out of the classes in S_i. The purpose is to better generalize the model and achieve high results on C_{novel} via training a different classification task for each episode. This approach is extended to object detection (FSOD) and instance segmentation (FSIS) [2], [4], [7], [45], [55]–[57]. Those proposals consider all objects in an image as queries, and they have a single support set per image instead of per query. However, FSIS is more challenging compared to few-shot classification or FSOD, as it has to consider semantic label, localization, and segmentation information. In our task, given a camouflage image I_q to query the model, FSIS returns labels y_i, bounding boxes b_i, and segmentation masks M_i for all camouflage objects in I_q that belong to the set of C_{test}.

B. Overview of Our FS-CAMOFreq Framework.

General Framework. Building on top of FS-CDIS [4], a pioneering work in few-shot camouflage object detection and instance segmentation, our proposed FS-CAMOFreq framework exploits the frequency-based data enhancement inspired by CamoFA [13] in an instance-aware manner. The ultimate goal is to provide more information to the model via the domain of frequency, where the global structural pattern and scale of variations are better expressed. We adopt the baseline of FS-CDIS architecture [4] with the version based on iFS-RCNN [2], which utilizes a two-stage training and fine-tuning scheme (illustrated in Figure 1). In the base phase, our model is trained on the COCO [58] dataset with abundant annotated data from 80 categories to achieve the base weights. Referenced from [4], the base weight reported from the ResNet-101 backbone trained on 80 classes of the COCO dataset yields the finest results among configurations. These weights are then utilized in the fine-tuning stage to continue to learn from novel camouflage categories from CAMO-FS [4] using a few samples of $K = \{1, 5\}$.

Few-Shot Fine-Tuning. Following the methodology of FS-CDIS [4] and iFS-RCNN [2], the input query images are processed through a feature extractor F, comprising a backbone network B, RoI align, RoI feature extraction modules, and a region proposal network. The model incorporates three specialized heads to support classification C, bounding box regression R, and mask prediction M. To this end, we employ the novel probit classifier for C and the uncertainty-guided

TABLE I
SOTA COMPARISON OF OUR FS-CAMOFREQ EVALUATED ON CAMO-FS [4]. THE UTILIZED BACKBONES ARE COCO-80 FPN-RESNET-101.

Model		nAP						nAP50						nAP75					
Method	Backbone/ Num. of shots	Instance Segmentation			Object Detection			Instance Segmentation			Object Detection			Instance Segmentation			Object Detection		
		1	5	Avg.	1	5	Avg.	1	5	Avg.	1	5	Avg.	1	5	Avg.	1	5	Avg.
MTFA [3]	COCO-80 ResNet-50	2.48	6.40	4.44	1.98	6.17	4.08	4.24	9.89	7.07	4.12	9.94	7.03	2.38	8.04	5.21	1.47	6.40	3.94
M-RCNN [28]		4.08	8.29	6.19	2.82	6.18	4.50	6.91	13.89	10.40	6.78	13.92	10.35	4.34	8.18	6.26	1.45	4.51	2.98
iFS-RCNN [2]		4.17	6.38	5.28	3.92	6.60	5.26	6.19	10.02	8.11	6.23	10.15	8.19	4.93	7.32	6.13	4.47	7.17	5.82
MTFA [3]	COCO-80 ResNet-101	3.66	5.95	4.81	2.93	5.84	4.39	5.37	8.67	7.02	5.86	9.13	7.50	4.09	6.94	5.52	2.20	6.04	4.12
M-RCNN [28]		4.39	10.09	7.24	3.03	7.79	5.41	7.58	15.41	11.50	7.53	15.86	11.70	4.53	11.90	8.22	1.42	5.34	3.38
iFS-RCNN [2]		4.27	7.80	6.04	3.79	8.08	5.94	5.98	11.35	8.67	5.92	11.52	8.72	4.75	9.15	6.95	4.46	9.24	6.85
FS-CDIS-ITL* [4]		5.35	9.35	7.35	4.71	10.36	7.54	7.80	14.01	10.91	7.85	14.40	11.13	6.04	11.57	8.81	5.51	11.32	8.42
FS-CDIS-IMS* [4]		2.99	9.03	6.01	2.74	8.44	5.59	4.62	12.48	8.55	4.81	13.18	9.00	3.36	9.82	6.59	2.98	9.69	6.34
Our performance																			
Baseline FS-CAMOFreq †	COCO-80 ResNet-101	5.55	8.21	6.88	5.34	8.82	7.08	8.42	12.07	10.25	8.49	12.86	10.68	6.19	9.58	7.89	5.98	9.22	7.60
FS-CAMOFreq (ours)		**5.71**	**8.31**	**7.01**	**5.56**	**8.89**	**7.23**	**8.50**	**11.72**	**10.11**	**8.56**	**12.11**	**10.34**	**6.46**	**9.53**	**8.00**	**6.25**	**9.49**	**7.87**

** denotes the FS-CDIS results built on top of iFS-RCNN [2]*
† denotes our reproduced baseline FS-CDIS iFS-RCNN [2], [4] on our upgraded CUDA version 12.4

bounding box predictor for R from the base model iFS-RCNN [2]. During the first training phase on base categories C_{base} from COCO, all components are jointly optimized. In the subsequent fine-tuning stage for novel categories of CAMO-FS, the backbone B is frozen to preserve learned representations, and only the three prediction heads C, R, and M are updated to adapt to new class instances. In our implementation of FS-CAMOFreq, we strictly adhere to this two-stage process to fine-tune on few-shot settings.

C. Instance-Aware Frequency-Based Data Enhancement.

In this work, we propose an instance-aware frequency domain enhancement strategy specifically tailored for camouflage images. Concretely, we apply frequency-based transformations selectively to the non-instance regions, leaving the target object unaltered. This targeted manipulation aims to amplify the visual contrast between the foreground instance and its background, thereby making the camouflaged object more distinguishable to the segmentation model. The below Equation 1, 2, and 3 well express the procedure.

Let $\mathcal{I}$ be the query image and $\mathcal{I}_R$ a reference image taken from the intra-class training set with $\mathcal{I}_R$. Let $M \in \{0,1\}^{H\times W}$ be the binary mask where $M_{i,j} = 1$ denotes foreground pixels and 0 otherwise. We first transform both images into the frequency domain via Fast Fourier Transform:

$$\mathcal{F}(\mathcal{I}) = \mathcal{A}(\mathcal{I}) \cdot e^{j\phi(\mathcal{I})}, \quad \mathcal{F}(\mathcal{I}_R) = \mathcal{A}(\mathcal{I}_R) \cdot e^{j\phi(\mathcal{I}_R)} \tag{1}$$

To perform instance-aware augmentation, we swap the amplitude of the background (where $M = 0$) with that of the reference image $\mathcal{I}_R$, while keeping the original foreground amplitude and phase unchanged:

$$\mathcal{A}_m = M \cdot \mathcal{A}(\mathcal{I}) + (1 - M) \cdot \mathcal{A}(\mathcal{I}_R) \tag{2}$$

Finally, the augmented image $\mathcal{I}_*$ is reconstructed via inverse Fast Fourier Transform iFFT:

$$\mathcal{I}_* = \mathcal{F}^{-1}(\mathcal{A}_m \cdot e^{j\phi(\mathcal{I})}) \tag{3}$$

resulting in a new image that retains the spatial structure of the query but exhibits enriched frequency characteristics derived from the reference image. This augmentation is performed in an *instance-aware* fashion, ensuring that the frequency blending process is applied selectively to regions guided by the instance mask, thus preserving semantic consistency while enhancing visual diversity.

IV. EXPERIMENTS

A. Experimental Configurations

Few-shot CIS Benchmark. In few-shot CIS, we observe a shortage of benchmark datasets established to evaluate the task. Following the work of Nguyen *et al.*, we evaluate our FS-CAMOFreq on the challenging CAMO-FS [4] dataset. CAMO-FS is proposed as the pioneer dataset in the camouflage domain to support multiple vision tasks, publishing labels in all classification, object detection, and instance segmentation tasks. It is prepared to evaluate few-shot learning with annotation in COCO JSON format. CAMO-FS includes $2,852$ images in total, providing 197 images (with 235 instances) for training few-shot learning and the remaining $2,655$ images (with $3,107$ instances) for testing.

Settings. Based on the published settings of the FSOD and FSIS methods [1]–[4], [55], [57], we set up the configurations for our FS-CAMOFreq framework. We use the recent baseline FS-CDIS [4] built on top of iFS-RCNN [2] to conduct our experiments. The implementation uses the base framework of Detectron2 [59], the backbone is the common ResNet-101 [60] with Feature Pyramid Network [61]. Basically, the models are trained in two stages: base phase and novel fine-tuning phase. In the base phase, we train the framework with abundant annotated data taken from 80 semantic classes with over $118K$ images in the training set of the generic COCO dataset [58]. This base training stage strictly follows the instructions of Detectron2 [59] regarding training configurations. In the novel phase, we fine-tune the model on $K = \{1, 5\}$ shots corresponding to each novel class. The learning rate in this stage is $lr = 0.01$ with batch size $bz = 16$ inferred from iFS-RCNN [2]. Other training hyperparameters of the novel phase are followed by FS-CDIS [4] settings. After training, the framework is evaluated on the test set of CAMO-FS, which includes $2,655$ images with $3,107$ instances of 47 camouflage classes, to obtain the final performance. Please visit [2], [4] or [59] for more details on other parameters of both the training and testing phases. Our models are trained and tested on a single GeForce RTX 3090 Ti GPU with CUDA version 12.4.

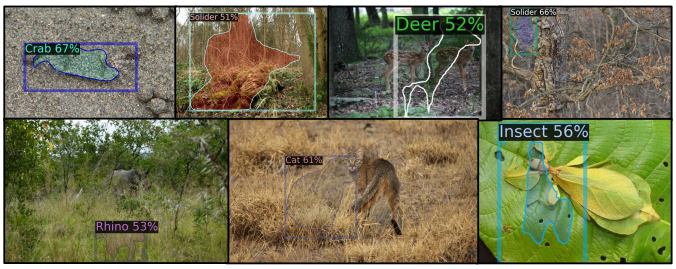

Fig. 2. Visualization results of our FS-CAMOFreq on the CAMO-FS [4]. The results are visualized under the configuration of the 5-shot setting, where the instance-aware frequency-based enhancement is already applied. Best viewed in color and zoomed in.

The Fast Fourier Transform (FFT) and the inverse version iFFF functions are implemented using the PyTorch FFT package.

Evaluation metrics. To report our object detection and instance segmentation results, we leverage COCO-based average precision (AP) and average recall (AR) metrics. In detail, we report AP@50 and AP@75, along with AR@10. We also report AP and AR at different scales of small, medium, and large. Please reach this site https://cocodataset.org/#detection-eval for details on the evaluation metrics.

B. State-of-the-art Few-shot CIS Comparison

To demonstrate the contribution of our FS-CAMOFreq, we establish experiments on the CAMO-FS benchmark [4] and compare our performance with other state-of-the-art (SoTA) methods in this few-shot approach, including Mask R-CNN [28], iMTFA [3], iFS-RCNN [2], and the baseline FS-CDIS [4]. We report the results of $K = \{1, 5\}$ shots to represent the extreme cases of one and few samples in CAMO-FS. By the way, we provide nAP, nAP50, and nAP75 on each result. We also leveraged the published source code of the aforementioned models to reproduce and report their results. Table I presents the evaluation of our FS-CAMOFreq among SoTA methods under different backbones. Among metrics, the nAP is the most important to express the effectiveness of a method, followed by nAP50 and nAP75. To this end, our baseline FS-CAMOFreq implemented on a newer compatible version of CUDA 12.4 (compared to the previous version published in [4]) yields higher accuracy of 5.55% on 1-shot and 8.21% on 5-shot instance segmentation; the respective values in object detection are 5.34% and 8.82%. Then we improve on top of the baseline, which is best described in nAP and nAP75 metrics. In detail, we achieve the following results measured by nAP, nAP50, and nAP75: $5.71\%, 8.50\%, 6.46\%$ in 1-shot instance segmentation; $8.31\%, 11.72\%, 9.53\%$ in 5-shot instance segmentation. The performance in object detection also increases by a large margin. The report demonstrates the effectiveness of our instance-aware frequency-based data enhancement method by providing information related to structural patterns and variations to the few-shot models.

Ablation Study on Instance Augmentation. In Table II, we present an ablation study where frequency-based augmentation is applied to the foreground region $M_{i,j} = 1$ instead of the background. This inverse setting, where foreground amplitudes are replaced with those from a reference image, results in performance degradation. The decline suggests that altering foreground features disrupts camouflage cues, leading to model confusion and reduced accuracy.

TABLE II
ABLATION STUDY OF OUR FS-CAMOFREQ ON INSTANCE REGION AUGMENTATION EVALUATED ON CAMO-FS [4].

FS-CAMOFreq	Detection			Segmentation		
Num. of shots	nAP	nAP50	nAP75	nAP	nAP50	nAP75
1	5.63	8.38	6.44	5.31	8.44	5.97
2	5.64	8.10	6.56	5.65	8.36	6.49
3	4.94	7.17	5.71	5.16	7.35	5.78
5	6.12	9.01	6.59	6.84	9.64	7.53
Avg.	**5.58**	**8.17**	**6.33**	**5.74**	**8.45**	**6.44**

Discussion. In Figure 2, we visualize the results of our FS-CAMOFreq on the 5-shot setting. Although the proposed method leads to improvements in nAP, the overall prediction quality remains limited. While the model successfully handles many cases, it still encounters failures such as missed detections, over-segmentation, and misclassification when the camouflage images appear to be difficult. With the reported results, we claim the impact of the frequency domain augmentation to the camouflage instance segmentation under a few-shot learning context. However, as observed in the work of [13], despite the improvement in all fully supervised configurations of different baseline methods, the absolute values of improvement are limited. Thus, the improvement when we extend the approach to a few-shot learning task, which is intense due to the limited samples, is also challenging.

V. CONCLUSION

In this paper, we address the challenging task of few-shot instance segmentation in camouflaged scenarios, where subtle differences between foreground and background limit the effectiveness of conventional approaches. While prior methods primarily rely on color space augmentations, we introduce FS-CAMOFreq, a novel framework that enriches image representations through instance-aware frequency domain augmentation. By leveraging the frequency spectrum information, our method enhances feature diversity without compromising semantic coherence, enabling better generalization under low-data camouflage constraints. Extensive experiments on the CAMO-FS benchmark validate the superiority of our approach over existing state-of-the-art baselines, highlighting the potential of frequency-based augmentations in advancing few-shot instance segmentation. Our future directions include exploring the adaptive frequency approach and applying our framework to other dense prediction tasks under limited supervision.

VI. ACKNOWLEDGEMENT

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number DS.C2025-26-08.

REFERENCES

[1] B.-B. Gao, X. Chen *et al.*, "Decoupling classifier for boosting few-shot object detection and instance segmentation," *NeurIPS*, vol. 35, pp. 18 640–18 652, 2022.

[2] K. Nguyen and S. Todorovic, "ifs-rcnn: An incremental few-shot instance segmenter," in *CVPR*, 2022, pp. 7010–7019.

[3] D. A. Ganea, B. Boom, and R. Poppe, "Incremental few-shot instance segmentation," in *CVPR*, June 2021, pp. 1185–1194.

[4] T.-D. Nguyen, A.-K. N. Vu, N.-D. Nguyen, V.-T. Nguyen, T. D. Ngo, T.-T. Do, M.-T. Tran, and T. V. Nguyen, "The art of camouflage: Few-shot learning for animal detection and segmentation," *IEEE Access*, 2024.

[5] M.-Q. Le, T. V. Nguyen *et al.*, "Maskdiff: Modeling mask distribution with diffusion probabilistic model for few-shot instance segmentation," in *AAAI*, vol. 38, no. 3, 2024, pp. 2874–2881.

[6] S. Chang, Y. Pang, X. Zhao, H. Lu, and L. Zhang, "Beyond mask: Rethinking guidance types in few-shot segmentation," *Pattern Recognition*, vol. 165, p. 111635, 2025.

[7] T.-D. Nguyen, V.-T. Nguyen, and T. V. Nguyen, "A generative approach at the instance-level for image segmentation under limited training data conditions (stu. abs.)," in *AAAI*, vol. 39, no. 28, 2025, pp. 29 451–29 452.

[8] Z. Wang, Y. Li, Y. Yang, Y. Li, and G. Liu, "Few-shot camouflaged object segmentation," in *IJCNN*. IEEE, 2024, pp. 1–10.

[9] T.-N. Le, V. Nguyen *et al.*, "Camoufinder: Finding camouflaged instances in images," in *AAAI*, 2021.

[10] T.-N. Le *et al.*, "Camouflaged instance segmentation in-the-wild: Dataset, method, and benchmark suite," *IEEE TIP*, vol. 31, 2022.

[11] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," in *CVPR*, 2020, pp. 2777–2787.

[12] A. Khan, M. Khan *et al.*, "Camofocus: Enhancing camouflage object detection with split-feature focal modulation and context refinement," in *WACV*, 2024, pp. 1434–1443.

[13] M.-Q. Le, M.-T. Tran, T.-N. Le, T. V. Nguyen, and T.-T. Do, "Camofa: A learnable fourier-based augmentation for camouflage segmentation," in *WACV*. IEEE, 2025, pp. 3427–3436.

[14] Y. Yang and S. Soatto, "Fda: Fourier domain adaptation for semantic segmentation," in *CVPR*, 2020, pp. 4085–4095.

[15] Y. B. Luo, J. H. Cai *et al.*, "Ffs-net: Fourier-based segmentation of colon cancer glands using frequency and spatial edge interaction," *Expert Systems with Applications*, vol. 262, p. 125527, 2025.

[16] P. Vaish, S. Wang, and N. Strisciuglio, "Fourier-basis functions to bridge augmentation gap: Rethinking frequency augmentation in image classification," in *CVPR*, 2024, pp. 17 763–17 772.

[17] X. Xu, J. Yang *et al.*, "Physaug: A physical-guided and frequency-based data augmentation for single-domain generalized object detection," in *AAAI*, vol. 39, no. 20, 2025, pp. 21 815–21 823.

[18] R. Azad, A. Bozorgpour, M. Asadi-Aghbolaghi, D. Merhof, and S. Escalera, "Deep frequency re-calibration u-net for medical image segmentation," in *ICCV*, 2021, pp. 3274–3283.

[19] B. Cheng, L.-C. Chen *et al.*, "Spgnet: Semantic prediction guidance for scene parsing," in *CVPR*, 2019.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015.

[21] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE TPAMI*, vol. 40, no. 4, pp. 834–848, 2017.

[22] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for semantic segmentation," in *ICCV*, 2021.

[23] E. Xie, W. Wang *et al.*, "Segformer: Simple and efficient design for semantic segmentation with transformers," in *NeurIPS*, 2021.

[24] J. Jain, A. Singh, N. Orlov, Z. Huang, J. Li, S. Walton, and H. Shi, "Semask: Semantically masked transformers for semantic segmentation," in *ICCV*, 2023, pp. 752–761.

[25] L. Zhang, W. Huang, and B. Fan, "Sarformer: Segmenting anything guided transformer for semantic segmentation," *Neurocomputing*, vol. 635, p. 129915, 2025.

[26] A. Dosovitskiy, L. Beyer *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *ICLR*, 2021.

[27] A. Vaswani, N. Shazeer *et al.*, "Attention is all you need," *NeurIPS*, vol. 30, 2017.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *ICCV*, 2017, pp. 2980–2988.

[29] Z. Cai and N. Vasconcelos, "Cascade r-cnn: High quality object detection and instance segmentation," *IEEE TPAMI*, pp. 1483–1498, 2019.

[30] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," in *ICCV*, 2019, pp. 9197–9206.

[31] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *CVPR*, 2018.

[32] S. Qiao, L.-C. Chen, and A. Yuille, "Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution," in *CVPR*, 2021, pp. 10 213–10 224.

[33] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "Yolact: Real-time instance segmentation," in *ICCV*, 2019.

[34] X. Wang, R. Zhang, C. Shen, T. Kong, and L. Li, "Solo: A simple framework for instance segmentation," *IEEE TPAMI*, vol. 44, no. 11, pp. 8587–8601, 2021.

[35] B. Cheng, I. Misra *et al.*, "Masked-attention mask transformer for universal image segmentation," in *CVPR*, 2022, pp. 1290–1299.

[36] J. He, P. Li, Y. Geng, and X. Xie, "Fastinst: A simple query-based model for real-time instance segmentation," in *CVPR*, 2023, pp. 23 663–23 672.

[37] J. Jain, J. Li, M. T. Chiu, A. Hassani, N. Orlov, and H. Shi, "Oneformer: One transformer to rule universal image segmentation," in *CVPR*, 2023.

[38] T.-N. Le, T. V. Nguyen, Z. Nie *et al.*, "Anabranch network for camouflaged object segmentation," *CVIU*, vol. 184, pp. 45–56, 2019.

[39] N. Dong and E. P. Xing, "Few-shot semantic segmentation with prototype learning." in *BMVC*, vol. 3, no. 4, 2018.

[40] W. Liu, C. Zhang, G. Lin, and F. Liu, "Crnet: Cross-reference networks for few-shot segmentation," in *CVPR (CVPR)*, June 2020.

[41] O. Saha, Z. Cheng, and S. Maji, "Ganorcon: Are generative models useful for few-shot segmentation?" in *CVPR*, June 2022.

[42] Z. Tian, X. Lai *et al.*, "Generalized few-shot semantic segmentation," in *CVPR*, June 2022, pp. 11 563–11 572.

[43] Y. Han, J. Zhang *et al.*, "Reference twice: A simple and unified baseline for few-shot instance segmentation," *IEEE TPAMI*, 2024.

[44] H. Wang, J. Liu *et al.*, "Dynamic transformer for few-shot instance segmentation," in *ACMMM*, 2022, pp. 2969–2977.

[45] B.-B. Gao, X. Chen, Z. Huang, C. Nie, J. Liu, J. Lai, G. Jiang, X. Wang, and C. Wang, "Decoupling classifier for boosting few-shot object detection and instance segmentation," in *NeurIPS*, 2022.

[46] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," *NeurIPS*, vol. 31, 2018.

[47] A. Santhirasekaram, M. Winkler, A. Rockall, and B. Glocker, "A geometric approach to robust medical image segmentation," *Medical Image Analysis*, vol. 97, p. 103260, 2024.

[48] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *Journal of medical imaging and radiation oncology*, vol. 65, no. 5, pp. 545–563, 2021.

[49] K. Wang, B. Fang, J. Qian, S. Yang, X. Zhou, and J. Zhou, "Perspective transformation data augmentation for object detection," *IEEE Access*, vol. 8, pp. 4935–4943, 2019.

[50] R. Niri, E. Gutierrez, H. Douzi, Y. Lucas, S. Treuillet, B. Castañeda, and I. Hernandez, "Multi-view data augmentation to improve wound segmentation on 3d surface model by deep learning," *IEEE Access*, vol. 9, pp. 157 628–157 638, 2021.

[51] L. Chen, L. Gu, and Y. Fu, "When semantic segmentation meets frequency aliasing," *arXiv preprint arXiv:2403.09065*, 2024.

[52] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *NeurIPS*, vol. 30, 2017.

[53] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra *et al.*, "Matching networks for one shot learning," *NeurIPS*, vol. 29, 2016.

[54] S. Gidaris and N. Komodakis, "Dynamic few-shot visual learning without forgetting," in *CVPR*, 2018, pp. 4367–4375.

[55] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell, "Few-shot object detection via feature reweighting," in *ICCV*, 2019.

[56] Z. Fan, J.-G. Yu *et al.*, "Fgn: Fully guided network for few-shot instance segmentation," in *CVPR*, 2020, pp. 9172–9181.

[57] X. Yan, Z. Chen *et al.*, "Meta r-cnn: Towards general solver for instance-level low-shot learning," in *ICCV*, 2019.

[58] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*. Springer, 2014, pp. 740–755.

[59] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

[60] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.

[61] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, 2017.