Dual-Agent Optimization framework for Cross-Domain Few-Shot Segmentation

Supplementary Material

A. More Details about the CD-FSS Benchmark

In this section, we provide additional details about the CD-FSS benchmark introduced in Section **Datasets and Evaluation Metrics** of the main paper. As mentioned, the benchmark comprises four datasets: Chest X-ray, ISIC, FSS-1000 and Deepglobe. Below, we elaborate on their characteristics:

- FSS-1000 [8] is a natural image dataset specifically designed for few-shot segmentation, featuring 1,000 object categories, each represented by 10 annotated samples. For our experiments, we adopt the official semantic segmentation split and evaluate models on the official test set, which includes 240 categories and a total of 2,400 images.
- DeepGlobe [4] is a satellite image dataset with pixel-level annotations for seven categories: urban, agriculture, rangeland, forest, water, barren, and unknown. Each image has a resolution of 2448 × 2448 pixels. Following the preprocessing in [7], we divide each image into six patches of 408 × 408 pixels to increase the number of test samples. After filtering out single-class images and the 'unknown' category, we evaluate models on 5,666 test images with their corresponding masks.
- ISIC2018 [3] is a dataset for skin lesion analysis, containing 2,596 annotated images, each with a single primary lesion. Following [7], we use the official training set for evaluation, resizing all images to a uniform resolution of 512×512 pixels from their original resolution of approximately 1022×767.
- Chest X-ray [1, 6] is a dataset for tuberculosis detection, consisting of 566 X-ray images (4020×4892 pixels) from 58 tuberculosis cases and 80 normal cases. Following [7], we downsize the images to 1024×1024 pixels for testing.

B. Detailed Experimental Settings

To ensure a fair comparison with [9], we adopt SSP [5] as our baseline. Consistent with SSP, we use the popular ResNet-50 pretrained on ImageNet as the backbone, omitting the final backbone stage and ReLU for improved generalization. Similar to [9], we integrate SSP's core components, including the adaptive self-support background prototype and self-support refinement modules. These enhancements effectively handle cluttered backgrounds and improve segmentation accuracy. All experiments were conducted under the same configurations to ensure consistency.

C. Model complexity

We conducted a comprehensive analysis of computational overhead in Table S1 across different methods. The results demonstrate that DATO achieves superior performance with only marginally increased computational costs compared to the state-of-the-art IFA method. Specifically, our method achieves notable performance gains to the SSP framework with minimal parametric overhead.

Table S1

Methods	#Params	FLOPs	FPS	mIoU
SSP	8.7M	27.78G	20.93	60.00
IFA	8.7M	28.01G	18.20	67.80
Ours	10.5M	30.12G	17.41	**70.31**

Table S2

Methods	Chest X-ray	ISIC	FSS-1000	Deepglobe	Average
APR	57.85	31.09	73.66	34.28	49.22
DR-Adapter	82.35	40.77	79.05	41.29	60.86
Ours	79.58	68.76	81.79	51.10	**70.31**

D. Discussion of related works in frequency and adversarial learning

We provide comprehensive empirical comparisons with state-of-the-art frequency-based methods (e.g., APM [10]) in Table 1 and additional frequency-based approaches (e.g., APR [2]) in Table S2, demonstrating our method's significant advantages. Notably, while previous adversarial learning approaches have primarily operated in the spatial domain, we innovatively combine adversarial learning with frequency domain analysis, which, to the best of our knowledge, is the first such application in the CD-FSS task.

E. More visualization results

Qualitative results in CD-FSS benchmark. As shown in Figure S1, the qualitative results demonstrate that our proposed method significantly enhances generalization performance under different levels of domain shift. Specifically, our approach effectively captures fine-grained details of the target objects and maintains robust segmentation performance even in challenging scenarios, such as drastic changes in lighting, texture, or background clutter. In contrast, baseline models often exhibit difficulties in these situations, frequently mis-segmenting the background regions or failing to fully activate the target object. This highlights the superior ability of our method to adapt across domains, avoiding common pitfalls such as overfitting to source domain characteristics or underperforming in unseen target domains. These results clearly validate the effectiveness of our method in addressing the domain shift problem and demonstrate its potential for real-world applications where target domains differ significantly from the training data.

Visual Comparison of Original and Agent-Enhanced Feature Representations. Figure S2 demonstrates the effectiveness of the proposed agents in enhancing feature representations. The agents first capture domain-invariant features from the data, which are then utilized to reinforce the

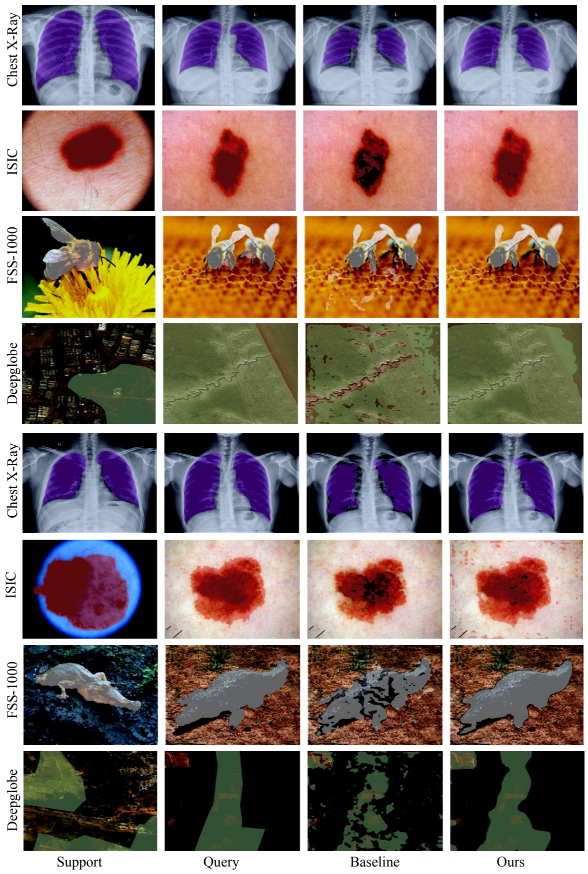

Figure S1. Qualitative results on the Chest X-ray, ISIC, FSS-1000, and Deepglobe datasets.

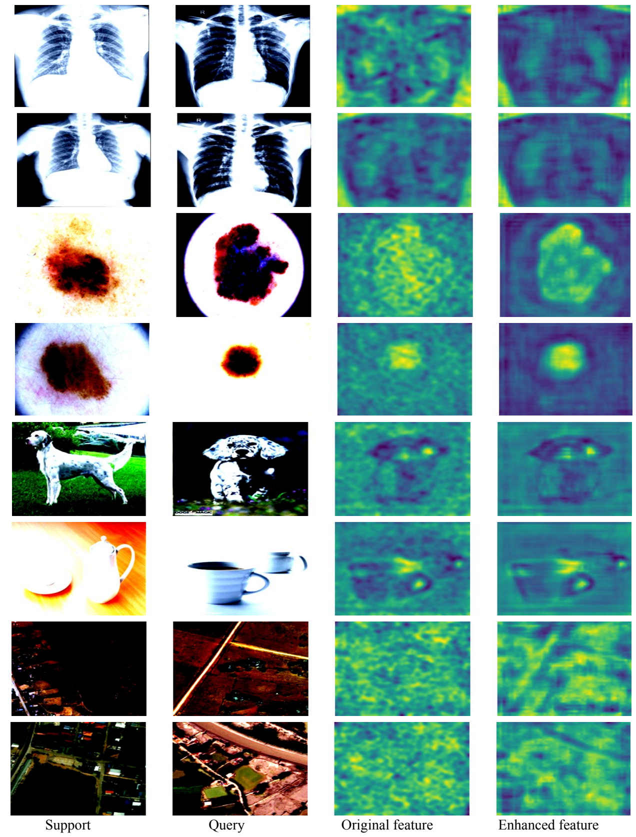

Figure S2. Visualization comparison between the original features and the agent-enhanced features demonstrates that the enhanced features exhibit reduced noise and more distinct contours.

domain-agnostic components within the original features. Simultaneously, the agents suppress domain-specific noise, resulting in cleaner and more distinct feature contours. This dual function not only reduces noise but also enhances the transferability of features across different domains, effectively mitigating the challenges posed by domain shifts.

References

[1] Sema Candemir, Stefan Jaeger, Kannappan Palaniappan, Jonathan P Musco, Rahul K Singh, Zhiyun Xue, Alexandros Karargyris, Sameer Antani, George Thoma, and Clement J McDonald. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE transactions on medical imaging*, 33(2):577–590, 2013. 1

[2] Guangyao Chen, Peixi Peng, Li Ma, Jia Li, Lin Du, and Yonghong Tian. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 458–467, 2021. 1

[3] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. 1

[4] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 172–181, 2018. 1

[5] Qi Fan, Wenjie Pei, Yu-Wing Tai, and Chi-Keung Tang. Self-support few-shot semantic segmentation. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, pages 701–719. Springer, 2022. 1

[6] Stefan Jaeger, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, Fiona Callaghan, Zhiyun Xue, Kannappan Palaniappan, Rahul K Singh, Sameer Antani, et al. Automatic tuberculosis screening using chest radiographs. *IEEE transactions on medical imaging*, 33(2):233–245, 2013. 1

[7] Shuo Lei, Xuchao Zhang, Jianfeng He, Fanglan Chen, Bowen Du, and Chang-Tien Lu. Cross-domain few-shot se-

mantic segmentation. In *European Conference on Computer Vision*, pages 73–90. Springer, 2022. 1

[8] Xiang Li, Tianhan Wei, Yau Pun Chen, Yu-Wing Tai, and Chi-Keung Tang. Fss-1000: A 1000-class dataset for few-shot segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2869–2878, 2020. 1

[9] Jiahao Nie, Yun Xing, Gongjie Zhang, Pei Yan, Aoran Xiao, Yap-Peng Tan, Alex C Kot, and Shijian Lu. Cross-domain few-shot segmentation via iterative support-query correspondence mining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3380–3390, 2024. 1

[10] Jintao Tong, Yixiong Zou, Yuhua Li, and Ruixuan Li. Lightweight frequency masker for cross-domain few-shot semantic segmentation. *arXiv preprint arXiv:2410.22135*, 2024. 1