Cross-Domain Few-Shot Segmentation via Iterative Support-Query Correspondence Mining

Supplementary Material

A1. Cross-Domain Few-Shot Segmentation

Motivation. Few-Shot Segmentation (FSS) methods rely on abundant base categories data to learn the capability of segmenting novel categories with a few exemplars [10, 13, 40, 43, 46, 47, 64]. However, collecting sufficient annotated data is infeasible in low-resource domains (*e.g.*, satellite images and medical screenings), thus the FSS pipeline is no longer suitable. Cross-Domain Few-Shot Segmentation (CD-FSS) [33] proposes a possible solution for the above challenge. Specifically, it aims to meta-train models on a source domain (*e.g.*, Pascal VOC [11]) with abundant accessible data, and adapt trained models to low-resource domains with a small support set (refer to Fig. A1). CD-FSS prposes an efficient solution for some specific downstream tasks (*e.g.* TB detection and wildlife conservation), where collecting substantial data is laborious, costly, and may raise privacy issues [33]. The CD-FSS pipeline eases the efforts of collecting and annotating large amounts of data in low-resource target domains.

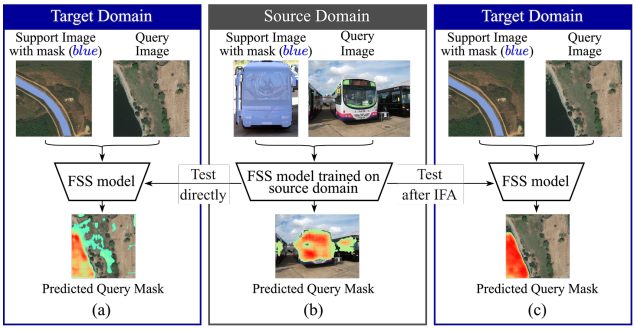

Figure A1. The formulation of Cross-Domain Few-Shot Segmentation task. (a) The segmentation performance clearly suffers from a severe drop when directly applying the trained model to target images. (b) Meta-training model on the source domain. (c) The segmentation performance is clearly improved benefiting from our proposed Iterative Few-shot Adaptor (IFA).

Comparison with Domain-Shift Few-Shot Segmentation. Although Boudiaf *et al*. [1] propose a Domain-Shift Few-Shot Segmentation (DS-FSS) setting and claim it is more realistic compared with FSS benchmarks by using data in a different domain for evaluation. Nevertheless, both base and novel categories are from daily object datasets. Therefore, the images in DS-FSS setting are easy to collect in large quantities, which is not a challenge in real application scenarios. Moreover, only utilizing daily object categories for evaluation can not exhibit the generalization capacity entirely. Consequently, Lei *et al*. [33] explore specific domains within CD-FSS task, which is more challenging: *(i)* satellite images and medical screenings are difficult to collect due to expensive cost and privacy agreements respectively, *(ii)* tiny or scarce objects are always neglected even in few-shot datasets [41, 44].

Existing problem. A feasible solution for CD-FSS is to perform meta-training on an annotation-abundant domain (*e.g.*, Pascal VOC [11]) and subsequently transfer to target data-limited ones. However, the learned models often suffer from a severe performance drop when directly applied to a different domain as shown in Fig. A1(a), and such a problem cannot be easily tackled by simply improving the few-shot capability. Because the core challenge is brought by the clear domain gap, we propose Iterative Few-shot Adaptor (IFA). In such a way, a model mines more support-query correspondence with extremely limited data, and generalized the capability of segmenting novel categories to target domains (refer to Fig. A1(c)).

A2. More Details of Applying SSP

We use Self-Support Prototype (SSP) [13] as our baseline, and also inherit its specific designs. To mitigate the effect form cluttered background, we incorporate the adaptive self-support background prototype [13] into our framework. Besides, we also use self-support refinement to achieve more accurate predictions, which is effective in [13]. For fair comparisons, we conduct all the experiments of the main paper with the same setup.

A3. Experimental Result

Results under Domain-Shift Few-Shot Segmentation. The concrete performances under Domain-Shift Few-Shot Segmentation (DS-FSS) are shown in Tab. A3 We observe that our IFA surpasses the previous best method with large margins in all folds, which validates the robustness and effectiveness of our designs.

More visualization results. We display more qualitative prediction results of our proposed Iterative Few-shot Adaptor (IFA), as shown in Fig. A3. It is obvious that IFA successfully transfers the capability learned from the source domain to four target domains. In Deepglobe [8], our IFA segments different categories from satellite images with similar accuracy. Besides, IFA segments medical screenings accurately from ISIC [7] and Chest X-Ray [49]. For

FSS-1000 [35], IFA predicts target objects across different types (*e.g.*, logos, foods, and objects) with satisfying results.

A4. Analysis

Effectiveness of iteration design. We also visualize the prediction after different iterations of BFP, as shown in Fig. A2. We observe that the iterative design increases the prediction results significantly. This validates our assumption outlined in the main paper that our IFA provides extensive information for mining support-query correspondence.

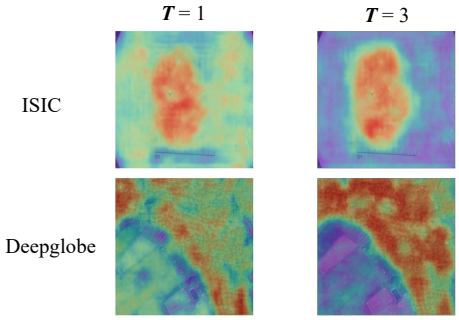

Figure A2. Visualized segmentation result (examples from ISIC and Deepglobe) after $\boldsymbol{T}$ times of recursive prediction.

Iterative computing cost. The iteration design is only used in fine-tuning, which only increases a little bit costs. Specifically, the time of each fine-tuning epoch increases to 12.87s from 4.82s, and the GPU memory occupies extra 29.3 Mb. The training and testing stages are unaffected.

Pascal VOC 2012 → Deepglobe									
$\lambda_{s'}$	0.1	0.2	0.3	**0.4**	0.5	0.6	0.7	0.8	0.9
mIoU	50.2	50.1	50.3	**50.6**	50.4	50.4	50.3	50.2	49.8
λ_i	**0.1**	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
mIoU	**50.6**	50.1	49.9	49.9	50.3	50.2	49.8	49.7	50.2
λ_{bs}	0.1	**0.2**	0.3	0.4	0.5	0.6	0.7	0.8	0.9
mIoU	50.4	**50.6**	50.5	50.1	49.9	50.2	50.3	50.2	50.5
λ_{bq}		0.5		**1.0**		1.5		2.0	
mIoU		50.0		**50.6**		49.6		49.7	

Table A1. Impact of the hyper-parameters $\lambda_{s'}$, λ_i, λ_{bs}, and λ_{bq} which denote the weight of $\mathcal{L}_{s'}$, $\mathcal{L}_i$, $\mathcal{L}_{bs}$, and $\mathcal{L}_{bq}$ in main paper.

Hyper-parameter value. We conduct experiments to determine the value of $\lambda_{s'}$, λ_i, λ_{bs}, and λ_{bq}, which balance the loss terms in main paper. Specifically, we first meta-train on Pascal VOC and use IFA to transfer the learned model to Deepglobe (1-shot setup with Res-50 backbone). From Tab. A1, we observe that the best performance is achieved when $\lambda_{s'} = 0.4$, $\lambda_i = 0.1$, $\lambda_{bs} = 0.2$, and $\lambda_{bq} = 1.0$ thus we determine the value of these hyper-parameters in our all experiments.

Validating Gestalt principle on four target datasets. Fan *et al*. [13] prove the Gestalt principle [30] existing in the daily object dataset by the cosine similarity statistics. Following their methods, we also conduct experiments in four target domains of the CD-FSS task. Statistic information is shown in Tab. A2. We can find that pixels belonging to the same object have much higher similarity than the cross-object pixels in all datasets, thus the Gestalt principle and SSP method are still effective. For the FSS-1000 dataset, we use all the images to compute the cosine similarity. For the remaining datasets, we randomly pick 200 images in each category for computation, except for picking 100 images of class 2 in ISIC due to data insufficiency.

ForeGround Pixels Similarity							
Deepglobe		ISIC		Chest X-Ray		FSS-1000	
cross	intra	cross	intra	cross	intra	cross	intra
0.497	**0.552**	0.512	**0.526**	0.528	**0.554**	0.494	**0.563**

Table A2. Cosine similarity for cross/intra-object pixels in four CD-FSS datasets.

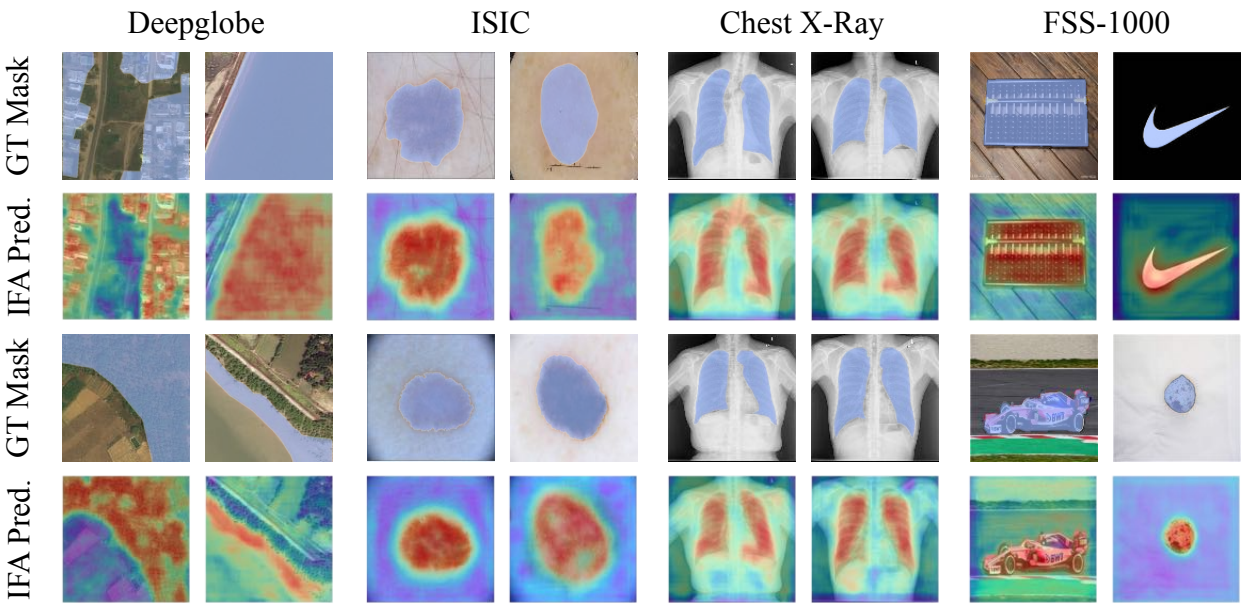

Figure A3. More qualitative results of Iterative Few-shot Adaptor (IFA) in four target datasets. Best viewed in color.

Source Domain: COCO-20i → Target Domain: Pascal-5i											
Methods	Backbone	1-shot					5-shot				
		fold-0	fold-1	fold-2	fold-3	**Mean**	fold-0	fold-1	fold-2	fold-3	**Mean**
RPMMs [59]	Res-50	36.3	55.0	52.5	54.6	49.6	40.2	58.0	55.2	61.8	53.8
PFENet [47]		-	-	-	-	60.8	-	-	-	-	61.9
RePRI [1]		52.4	64.3	65.3	71.5	63.3	57.0	68.0	70.4	76.2	67.9
ASGNet [34]		42.5	58.7	65.5	63.0	57.4	53.7	69.8	67.1	75.9	66.6
HSNet [40]		48.7	61.5	63.0	72.8	61.5	58.2	65.9	71.8	77.9	68.4
CWT [37]		53.5	59.2	60.2	64.9	59.4	60.3	65.8	67.1	72.8	66.5
Meta-Memory [54]		57.4	62.2	68.0	74.8	65.6	65.7	69.2	70.8	75.0	70.1
Ours$_{T=3}$		**61.9**	**71.4**	**68.7**	**82.0**	**71.0**	**73.2**	**82.1**	**80.4**	**88.0**	**80.9**
SCL [60]	Res-101	43.1	60.3	66.1	68.1	59.4	43.3	61.2	66.5	70.4	60.3
HSNet [40]		46.3	64.7	67.7	74.2	63.2	59.1	69.0	73.4	78.7	70.0
Meta-Memory [54]		59.4	64.3	70.8	72.0	66.6	67.2	72.7	72.0	78.9	72.7
Ours$_{T=3}$		**71.3**	**77.1**	**80.0**	**89.8**	**79.6**	**77.7**	**84.6**	**80.3**	**90.8**	**83.4**

Table A3. More detailed Quantitative comparison results on Domain-Shift Few-Shot Segmentation problem using mIoU (%) evaluation metric. The best and second best results are highlighted with **bold** and underline, respectively.