APSeg: Auto-Prompt Network for Cross-Domain Few-Shot Semantic Segmentation

Weizhao He[1†], Yang Zhang[1†*], Wei Zhuo[3*], Linlin Shen[1,2,3], Jiaqi Yang[3,4], Songhe Deng[1], Liang Sun[1]
[1]Computer Vision Institute, School of Computer Science & Software Engineering, Shenzhen University
[2]Shenzhen Institute of Artificial Intelligence and Robotics for Society
[3]National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University
[4]School of Computer Science, University of Nottingham, China
{heweizhao, dengsonghe, sunliang}2022@email.szu.edu.cn {yangzhang, weizhuo, llshen}@szu.edu.cn jiaqi.yang2@nottingham.edu.cn

1. Experiment

1.1. Datasets

FSS-1000. FSS-1000 [9] is a natural scenario dataset containing 1000 class categories with 10 samples per category. The evaluation procedure is conducted on 2400 randomly sampled support-query pairs.
Chest X-ray. Chest X-ray [1, 7] is an X-ray image dataset of 566 images collected from 58 cases with manifestations of tuberculosis and 80 normal cases.
ISIC. ISIC [4] is a skin lesion image dataset from the ISIC-2018 challenge. Following the previous approach [8], the evaluation procedure is conducted on the training set, which includes 2596 images and the corresponding annotations.
Deepglobe. Deepglobe [5] is a remote sensing image dataset that can be used for land cover segmentation. The dataset contains 7 categories: areas of urban, agriculture, rangeland, forest, water, barrel, and unknown. Following the previous method [8], we filter the unknown class in the training set and chunk the images to obtain 5666 images. We report the test results on the processed training set.

1.2. Implementation Details

In the meta prompt generator (MPG) module, the spatial size of the feature enrichment module (FEM) is set to $\{60, 30, 15, 8\}$, maintaining consistency with PFENet [10]. The transformer decoder [3] block consists of a self-attention mechanism, a cross-attention mechanism, and a feed-forward network. Its configuration is in line with Protoformer [2].

1.3. Ablation Study

Effect of the cycle consistent selection. In the dual prototype anchor transformation (DPAT) module, pseudo query prototypes are extracted through cycle-consistent selection (CSS) to enhance the feature transformation process. To further validate the effectiveness of CCS, we explore an alternative method for extracting pseudo query prototypes. Analogous to ResNet[6], we first obtain the coarse prediction mask of the query image and then perform the MAP operation to obtain query prototypes. This method is referred to as prediction mask-based MAP (PM-MAP). We conduct an experiment to evaluate the model without CCS, with CCS, and with PM-MAP, respectively, to better analyze the contribution of our CCS. The results in Tab. 1 show that CCS achieves better performance, with an average improvement of 2.67% mIoU on four datasets on the 1-shot setting. This indicates that CCS can extract reliable query prototypes to enhance the support prototypes, allowing features to be transformed into a more stable domain-agnostic space. Qualitative results on the effectiveness of CSS are provided in Fig. 1.

Method	1-shot mIoU
w/o CCS	58.63
w/ CCS	61.30
w/ PM-MAP	59.05

Table 1. Ablation studies of cycle-consistent selection (CSS) in dual prototype anchor transformation (DPAT) module. PM-MAP means mask-based MAP method. The results are averaged over 4 datasets under the 1-shot setting.

1.4. Additional Analysis

PerSAM [11] is a few-shot segmentation method based on SAM. To enable automatic segmentation, PerSAM extracts point prompts based on cosine similarity measure and it obtains box/mask prompts from the coarse predictions. The final prediction results are produced under the guidance of three types of visual prompts. The performance of PerSAM is far inferior to our proposed APSeg and PATNet [8]

[†] Equal Contribution: Weizhao He and Yang Zhang
[*] Corresponding Author: Yang Zhang and Wei Zhuo

due to the inability to extract precise prompts. In contrast, our APSeg introduces MPG and DPAT, avoiding reliance on precise visual prompts and achieving competitive results in cross-domain scenarios. Qualitative results are provided in Fig. 2. It can be observed that our method outperforms PerSAM by a large margin, which validates the effectiveness of our automatic way of generating prompt embeddings. In addition, we provide more qualitative segmentation results of our proposed method on four datasets in Fig. 3.

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] Leilei Cao, Yibo Guo, Ye Yuan, and Qiangguo Jin. Prototype as query for few shot semantic segmentation. *arXiv preprint arXiv:2211.14764*, 2022. 1

[3] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 1

[4] 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

[5] 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

[6] Xinyang Huang, Chuang Zhu, and Wenkai Chen. Restnet: Boosting cross-domain few-shot segmentation with residual transformation network. *arXiv preprint arXiv:2308.13469*, 2023. 1

[7] 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

[8] Shuo Lei, Xuchao Zhang, Jianfeng He, Fanglan Chen, Bowen Du, and Chang-Tien Lu. Cross-domain few-shot semantic segmentation. In *European Conference on Computer Vision*, pages 73–90. Springer, 2022. 1

[9] 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

[10] Zhuotao Tian, Hengshuang Zhao, Michelle Shu, Zhicheng Yang, Ruiyu Li, and Jiaya Jia. Prior guided feature enrichment network for few-shot segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 44(2):1050–1065, 2020. 1

[11] Renrui Zhang, Zhengkai Jiang, Ziyu Guo, Shilin Yan, Junting Pan, Hao Dong, Peng Gao, and Hongsheng Li. Personalize segment anything model with one shot. *arXiv preprint arXiv:2305.03048*, 2023. 1

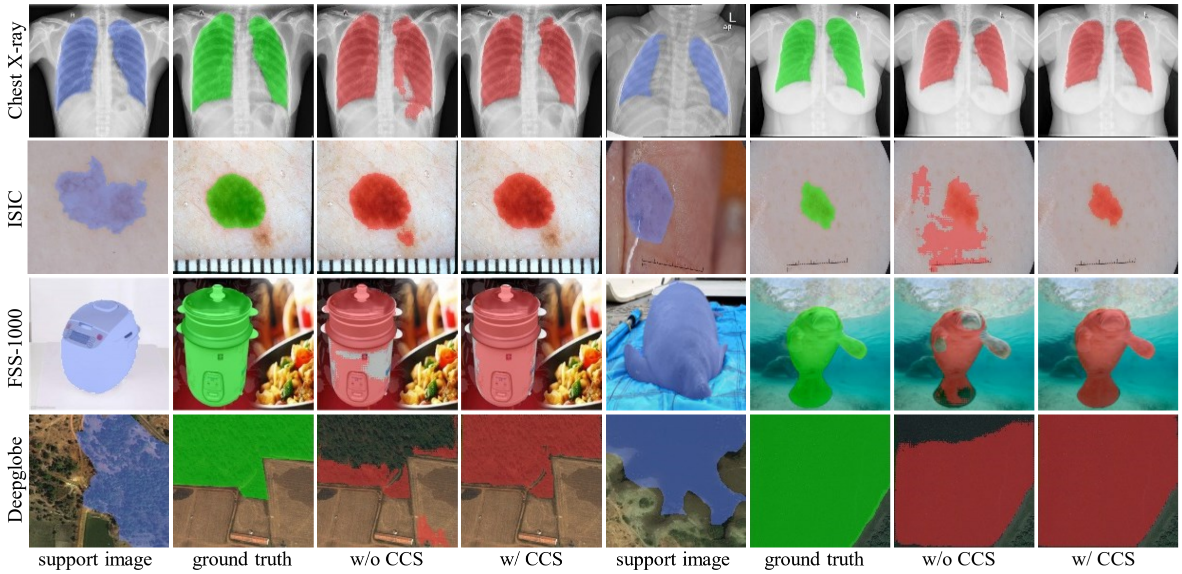

Figure 1. Visual comparison of segmentation results with and without cycle-consistent selection (CCS) in dual prototype anchor transformation (DPAT) module under the 1-shot setting.

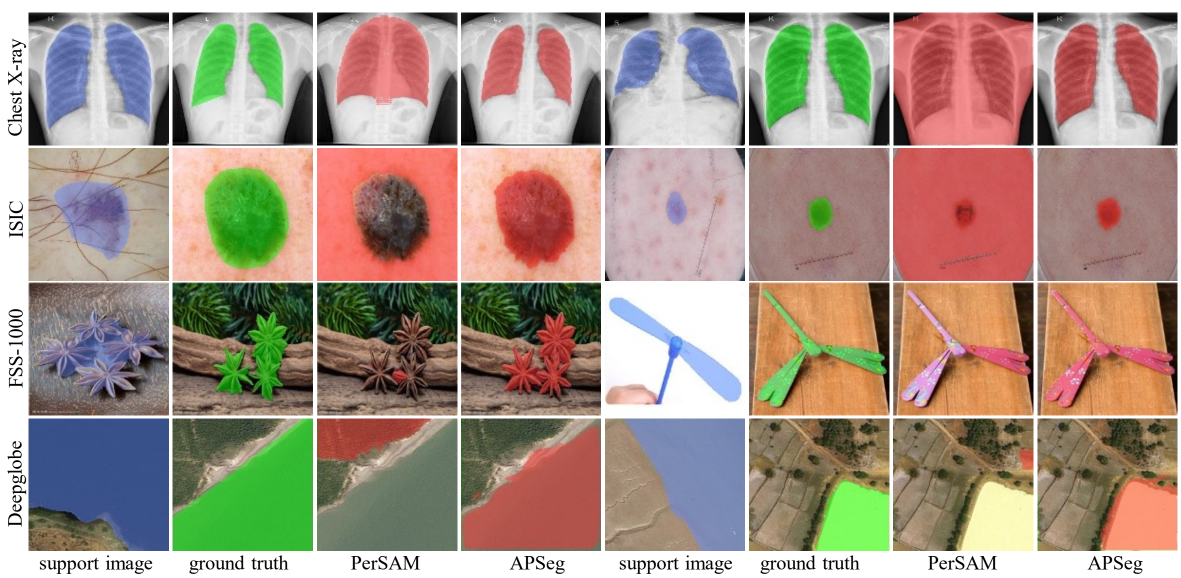

Figure 2. Visual Comparison Results between APSeg and PerSAM in four target datasets under the 1-shot setting.

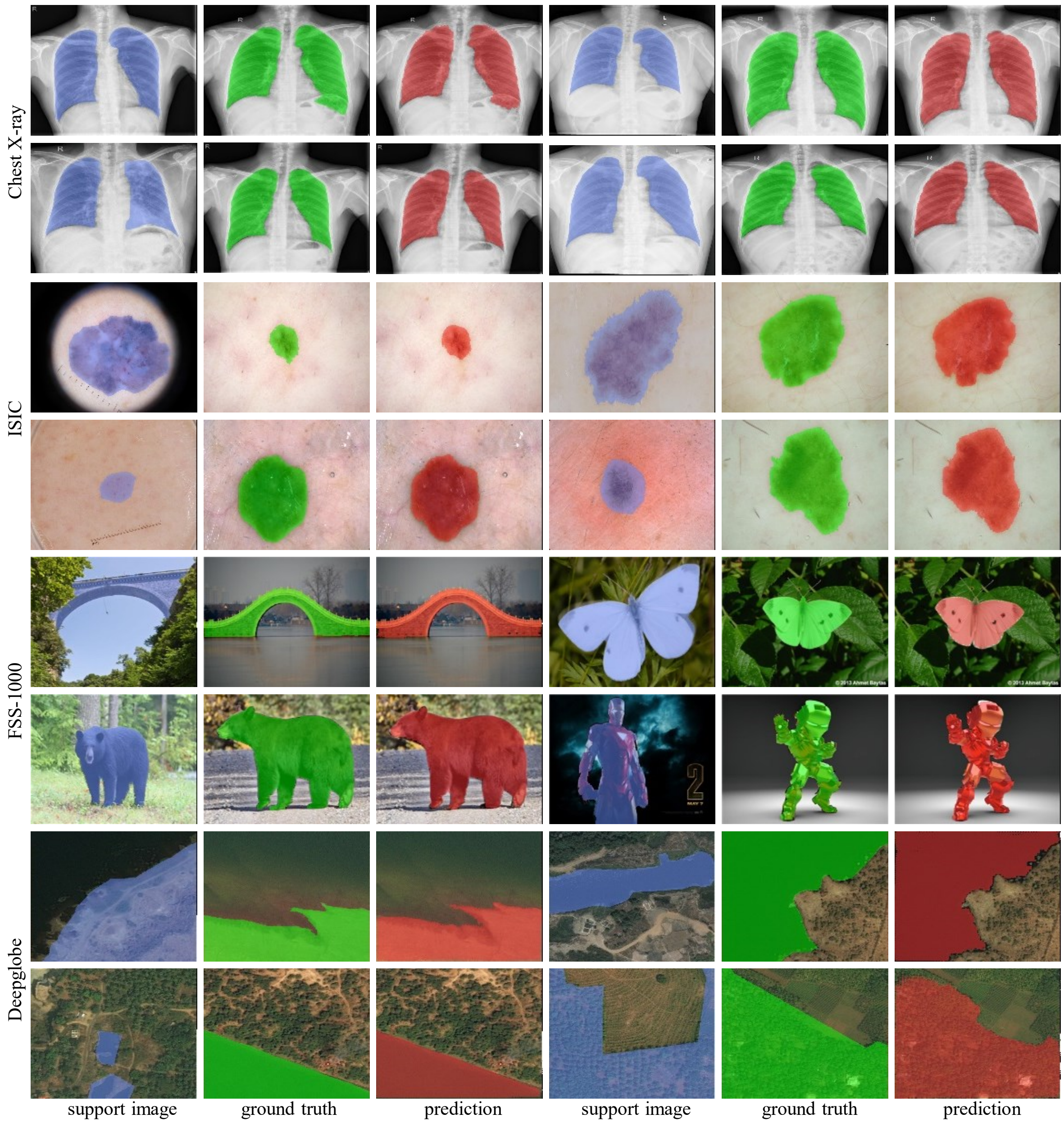

Figure 3. More qualitative segmentation results of our proposed APSeg in four target datasets under the 1-shot setting.