

Supplementary Materials: Anatomical Prior Guided Spatial Contrastive Learning for Few-Shot Medical Image Segmentation

Wendong Huang

Chongqing Key Laboratory of Image Cognition, Key
Laboratory of Cyberspace Big Data Intelligent Security,
Ministry of Education, Chongqing University of Posts and
Telecommunications
Chongqing, China
D220201013@stu.cqupt.edu.cn

Xiuli Bi

Chongqing Key Laboratory of Image Cognition, Key
Laboratory of Cyberspace Big Data Intelligent Security,
Ministry of Education, Chongqing University of Posts and
Telecommunications
Chongqing, China
bixl@cqupt.edu.cn

Jinwu Hu

Pazhou Lab, School of Software Engineering, South China
University of Technology
Guangzhou, China
202310189376@mail.scut.edu.cn

Bin Xiao*

Chongqing Key Laboratory of Image Cognition, Key
Laboratory of Cyberspace Big Data Intelligent Security,
Ministry of Education, Chongqing University of Posts and
Telecommunications
Chongqing, China
xiaobin@cqupt.edu.cn

In the supplementary material, additional details that are not included in the main paper due to space limits are provided as follows:

- Implementation details of the proposed anatomical prior guided spatial contrastive learning (APSCL).
- Details of additional experimental results.

1 MORE IMPLEMENTATION DETAILS

To evaluate the performance of the proposed 2D segmentation model on 3D organ scans, the evaluation protocol in prior methods [3, 4] is followed, where each 3D scan is reconverted to 2D slices and rescaled to the spatial resolution of 256×256 pixels for training and inference. Common conventions are followed in the data preprocessing stages, namely, random rotation, translation, and scaling. In addition, to tailor the pre-trained backbone, each 2D slice is duplicated three times on channel dimension. The training and testing pipelines for our method are illustrated in Fig. 1.

2 ADDITIONAL EXPERIMENTAL RESULTS

2.1 More Ablative Results

Effect of different backbones. To verify the effect of various backbones, we conduct ablation experiments on three popular backbones, namely VGG-16 [5], ResNet-50, and ResNet-101 [1]. Ablation results are summarized in Table 1, where these networks are all pre-trained on MS-COCO [2]. As can be noticed, in the 1-way 1-shot segmentation task, the model with ResNet-101 backbone has achieved the best performance of 85.47% in terms of the mean Dice Score on the CHAOS-T2 dataset, outperforming the model with the VGG-16 backbone and the model with the ResNet-50 backbone by 7.53% and 4.63%, respectively. The reason is that ResNet-101 with deeper layers is able to learn more discriminative features than other backbones. This indicates that using ResNet-101 as the backbone is conducive to achieving higher segmentation performance. Therefore, we choose ResNet-101 as our backbone.

**Corresponding authors

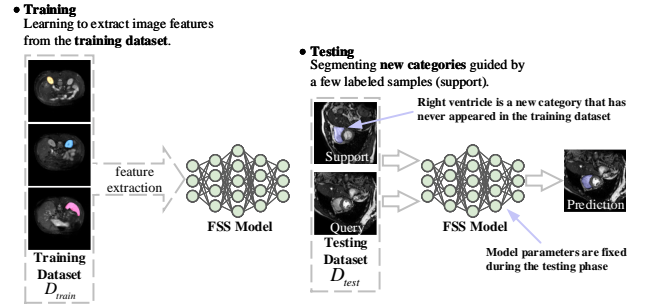


Figure 1: Paradigm of few-shot semantic segmentation (FSS).

Table 1: Ablation results (in Dice score %) using different backbones on the CHAOS-T2 dataset. The bold number denotes the best segmentation result.

Backbone	Liver	LK	RK	Spleen	Mean
VGG-16	78.72	77.37	81.39	74.26	77.94
ResNet-50	80.72	80.07	83.80	78.77	80.84
ResNet-101	86.73	84.66	89.66	80.82	85.47

Table 2: Ablative results (in Dice score %) of different KL losses on the CHAOS-T2 dataset.

Method	Liver	LK	RK	Spleen	Mean
\mathcal{L}_{KL}^s	84.07	82.01	87.58	78.20	82.97
\mathcal{L}_{KL}^q	83.29	83.74	86.03	78.49	82.89
$\mathcal{L}_{KL}^s + \mathcal{L}_{KL}^q$	86.73	84.66	89.66	80.82	85.47

Effect of the combined KL divergence. To verify the effectiveness of our combined KL divergence, *i.e.*, $\lambda_{KL} + \lambda_{SCL}$, we train

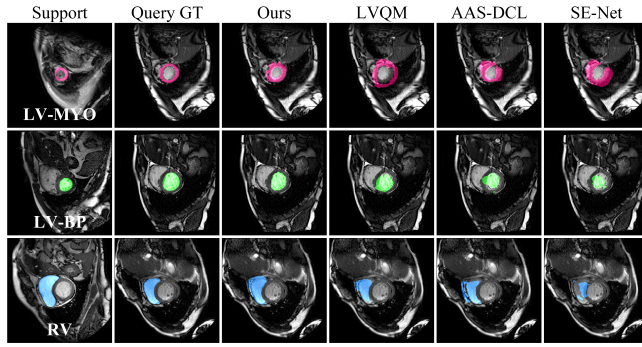


Figure 2: Visual comparison of the proposed method with other methods on the MS-CMRSeg dataset. GT is the ground truth.

our APSCL with λ_{KL} , λ_{SCL} , and $\lambda_{KL} + \lambda_{SCL}$, respectively. The segmentation results are presented in Table 2. As illustrated in Table 2, our combined KL divergence is more beneficial for anatomical prior learning.

2.2 More Qualitative Results

We compare the proposed APSCL and the current state-of-the-art methods in the 1-way 1-shot setting on the remaining medical imaging dataset: MS-CMRSeg. Fig. 2 present the performance comparison between our proposed APSCL and other methods. It is observed that the segmentation results of other methods on multiple organs all exist incomplete boundaries and inconsistent shapes, especially for SE-Net, while our method is resistant to interference from complex backgrounds and accurately predicts segmentation masks for query images.

2.3 Testing with Weak Annotations

We further assess our APSCL using the support set with weak annotations, including bounding boxes and scribbles. In the inference stage, dense pixel-wise annotations in the support set are replaced by bounding boxes or scribbles that are substituted from the original dense annotated masks automatically. In particular, bounding boxes refer to rectangular regions and scribbles refer to line-like regions. As shown in Table 3, our APSCL performs well when relying only on weak annotations and is capable of resisting the noise introduced by bounding boxes or scribbles. The performance with two different weak annotations is comparable to that using costly pixel-wise annotations, but the results using scribbles are on average 2.54%, 1.18%, and 2.98% higher than those using bounding boxes on the three medical datasets in terms of Dice scores, respectively. This is likely because scribbles provide more representative class information, whereas bounding boxes tend to bring more noise.

For a more intuitive comparison, we visualize the experimental results of different methods with various weak annotations in Fig. 3. Note that given different weak annotations, the model is still capable of to reasonably recognise the rough boundaries and shapes of new anatomical structures. This indicates the desirable ability of our method to tackle FSS tasks in complex scenarios.

Table 3: Quantitative results (in mean Dice score %) with different types of support annotations at inference time.

Annotation	CHAOS-T2	MS-CMRSeg	Synapse
Pixel-wise labels	85.47	77.99	81.50
Bounding boxes	81.20	74.36	77.26
Scribbles	83.74	75.54	80.24

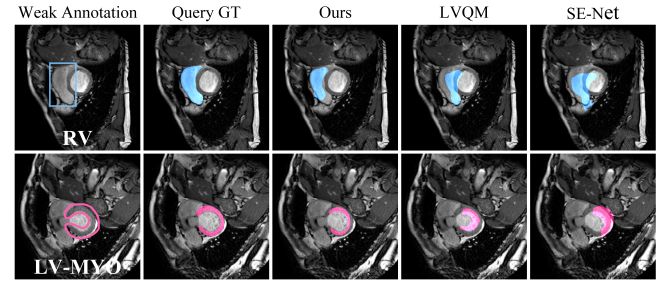


Figure 3: Qualitative results of the proposed APSCL and other methods on the MS-CMRSeg dataset with bounding box and scribble annotations at inference time. GT indicates the ground truth.

REFERENCES

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*. Springer, 740–755.
- [3] Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert. 2020. Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*. Springer, 762–780.
- [4] Abhijit Guha Roy, Shayan Siddiqui, Sebastian Pölsterl, Nassir Navab, and Christian Wachinger. 2020. ‘Squeeze & excite’ guided few-shot segmentation of volumetric images. *Medical image analysis* 59 (2020), 101587.
- [5] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).