Supplementary Material: Dual Contrastive Learning with Anatomical Auxiliary Supervision for Few-shot Medical Image Segmentation

Huisi Wu(✉), Fangyan Xiao, and Chongxin Liang

College of Computer Science and Software Engineering, Shenzhen University
hswu@szu.edu.cn, {2070276124, 2060271074}@email.szu.edu.cn

1 Non-target Slices as Motivation

For the *non-target slice set* defined in the paper, the images are derived from those slices in CT or MRI scans whose ground truths are all black. As shown in Fig. 1, these non-target slices contain rich anatomical knowledge, which can be utilized as background information to help the contrastive learning and guide the segmentation of the query image in the few-shot tasks.

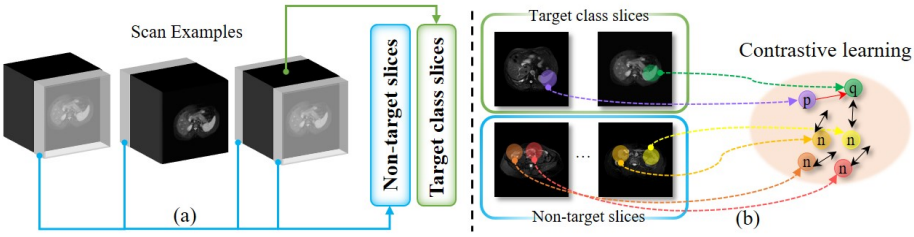

Fig. 1. (a) Illustrations of the non-target slices and target slices. (b) The motivation of utilizing non-target slices for contrastive learning

2 Detailed Process of the CIP Module

In order to show the operation process of the *constrained iterative prediction* (CIP) module more concretely, we further provide the corresponding algorithm, as shown in Algorithm 1.

3 More Experimental Results

3.1 More statistical results

Firstly, we show more numerical results to evaluate the impacts of different **balance parameters** for the contrastive losses in the total loss function (Equ.

Algorithm 1 Constrained Iterative Prediction

Input:

The query feature $\boldsymbol{f}^q$, the support feature $\boldsymbol{f}^s$, the support mask $\mathbf{y}^s$, the encoder features $\{\boldsymbol{f}_e^q\}$ and $\{\boldsymbol{f}_e^s\}$

Output:

The newest predicted query mask p_{new}^q

Given the query feature $\boldsymbol{f}^q$ of the final layer in the decoder and other information ($\boldsymbol{f}^s, \mathbf{y}^s, \{\boldsymbol{f}_e^q\}$, and $\{\boldsymbol{f}_e^s\}$), obtain the optimized prediction mask p_{new}^q
I = number of the total iteration, i = index of I
Employ the generic classifier (1×1 convolution and *softmax*) to generate the initial query prediction $p_i^q, i = 0$
Initialize $p_{new}^q = p_i^q, i = 0$
Initialize the ***similarity consistency constraint*** **(SCC)** as follows:
Compute the similarity map S_m between $\mathbf{y}^s$ and $\boldsymbol{p}_i^q, i = 0$; compute the similarity map S_f between $\{\boldsymbol{f}_e^q\}$ and $\{\boldsymbol{f}_e^s\}$; calculate the MSE loss between S_m and S_f to get initial loss $\mathcal{L}_{MSE}^i, i = 0$
Initialize $\mathcal{L}_{MSE}^{min} = \mathcal{L}_{MSE}^i, i = 0$
for all $i = 1, 2, ..., I$ **do**
    Enter the ***prediction head*** **(PredHead)**:
    Multiply the $\boldsymbol{f}^q$ and p_{new}^q to get $\boldsymbol{f}_m^q$; Send $\boldsymbol{f}^s$, $\mathbf{y}^s$ and $\boldsymbol{f}_m^q$ to *prior embedding* module to get $\hat{\boldsymbol{f}}_m^q$; The $\hat{\boldsymbol{f}}_m^q$ is applied to a series of convolutions and the generic classifier to obtain a new prediction mask p_i^q
    Enter the **SCC** module:
    Compute the similarity map S_m between $\mathbf{y}^s$ and $\boldsymbol{p}_i^q$; compute the similarity map S_f between $\{\boldsymbol{f}_e^q\}$ and $\{\boldsymbol{f}_e^s\}$; calculate the MSE loss between S_m and S_f to get the current MSE loss $\mathcal{L}_{MSE}^i$
    if $\mathcal{L}_{MSE}^i < \mathcal{L}_{MSE}^{min}$ **then**
        Update the newest predicted mask $p_{new}^q = p_i^q$
        Update $\mathcal{L}_{MSE}^{min} = \mathcal{L}_{MSE}^i$
    else
        Continue
    end if
end for
return p_{new}^q

6), as shown in Table 1. In addition, the Table 2 also demonstrates the effect of different iteration numbers of CIP module on the final segmentation results. And computational cost and parameters of models with different iterations in CIP are also shown in Table. 2. Finally, To evaluate if our improvement is statistically significant, we additionally conducted a Wilcoxon rank-sum test to compare our method with some of our competitors. From the p-values shown in Tables 3, we can clearly see that our method has a statistically significant improvement in terms of the Dice metric at the 5% level (all p-values are less than 0.05).

Table 1. Ablation study results (in Dice score) on the CHAOS-T2 dataset for balance parameters of contrastive losses

λ_{cpcl}	λ_{ppcl}	λ_{ccl}	Liver(%)	RK(%)	LK(%)	Spleen(%)	Mean(%)
0.10	0.10	0.08	61.39	74.27	68.84	70.23	68.68
0.10	0.08	0.06	64.61	75.49	68.42	72.87	70.35
0.08	0.14	0.03	66.67	79.72	70.35	74.60	72.84
0.08	0.12	0.03	69.65	78.07	72.89	73.39	73.50
0.08	0.12	0.04	67.54	82.16	72.34	73.22	73.82
0.08	0.12	0.06	**69.94**	**83.75**	**76.90**	**74.86**	**76.36**

Table 2. Ablation study results (in Dice score, computation costs and parameter numbers) on the CHAOS-T2 dataset for different iteration numbers (denoted as **I**) of the CIP module

I	Liver(%)	RK(%)	LK(%)	Spleen(%)	Mean(%)	Flops(G)	Params(M)
0	56.91	75.23	67.19	69.52	67.21	149.40	35.04
1	56.94	75.54	69.57	69.54	67.90	157.03	35.04
2	61.08	76.77	70.34	69.88	69.52	164.12	35.04
3	67.53	78.47	72.39	71.93	72.58	171.50	35.04
4	68.62	81.18	75.88	72.50	74.55	178.89	35.04
5	69.94	**83.75**	**76.90**	74.86	**76.36**	186.27	35.04
6	70.03	83.36	76.29	74.59	76.07	193.66	35.04
7	**70.11**	83.58	76.17	74.62	76.12	201.05	35.04
8	69.85	83.50	76.52	**74.99**	76.22	208.43	35.04
9	69.87	83.73	76.73	74.81	76.29	215.21	35.04

3.2 More visualized results

For the proposed CIP and DCL (including CCL and PCL) modules, more visualized ablation comparisons on the CHAOS-T2 dataset are shown in Fig. 2. Besides, we also present more visualized comparisons with classic and state-of-the-art methods (sSENet, SSL-ALPNet, and RP-Net), as shown in Fig. 3 and Fig. 4.

Table 3. Statistical significance analysis results (p-values on the Dice metric) on the CHAOS-T2 dataset for different methods. pv: p-value

Pairs	Liver(pv)	RK(pv)	LK(pv)	Spleen(pv)	Mean(pv)
PANet vs. Ours	2.53×10^{-7}	3.87×10^{-8}	5.79×10^{-8}	7.13×10^{-7}	2.85×10^{-8}
sSENet vs. Ours	1.82×10^{-8}	2.88×10^{-7}	3.87×10^{-7}	8.25×10^{-7}	2.74×10^{-7}
SSL-ALPNet vs. Ours	1.30×10^{-4}	5.63×10^{-4}	3.69×10^{-4}	1.47×10^{-5}	5.60×10^{-4}
RP-Net vs. Ours	2.19×10^{-4}	9.04×10^{-3}	8.17×10^{-3}	9.98×10^{-3}	1.01×10^{-2}

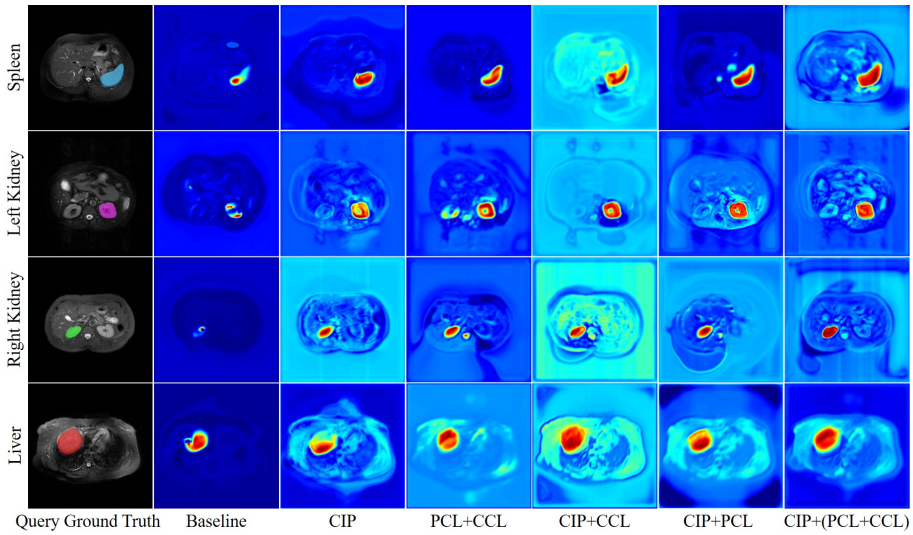

Fig. 2. More visual feature maps of the CIP and DCL modules on the CHAOS-T2 dataset. The warmer colors represent the better discriminative features. **DCL**: CCL+PCL; **PCL**: CPCL+PPCL

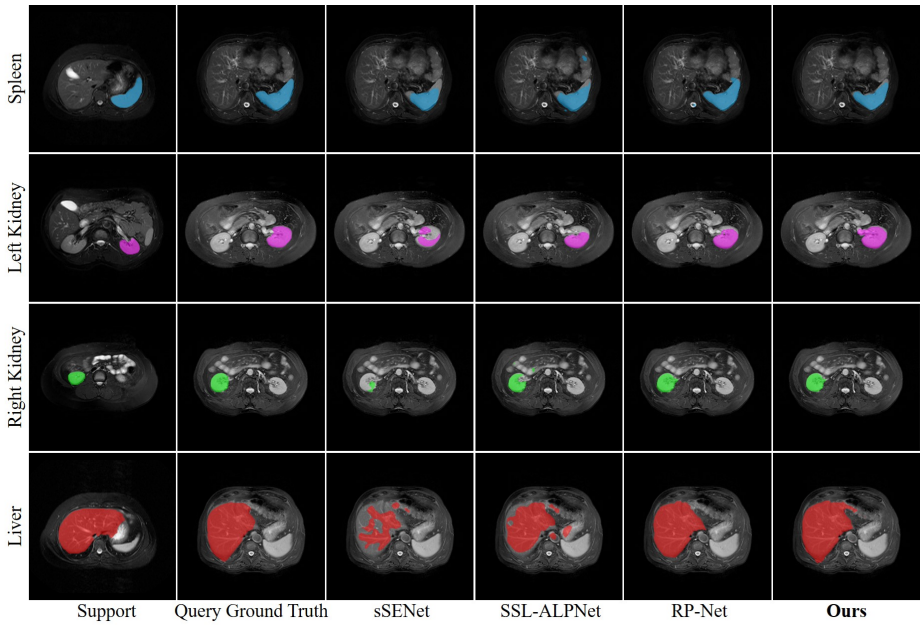

Fig. 3. More visual comparisons with different methods on the CHAOS-T2 dataset

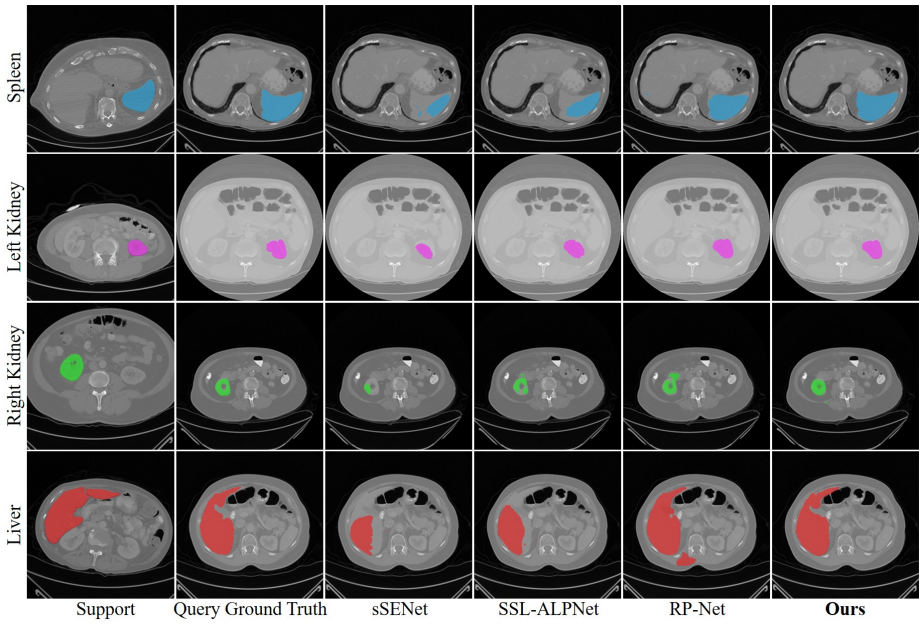

Fig. 4. More visual comparisons with different methods on the Synapse dataset