Supplementary Material: Exemplar Learning for Medical Image Segmentation

Qing En[1]
QingEn@cunet.carleton.ca
Yuhong Guo[1,2]
YuhongGuo@carleton.ca

[1] Carleton University
Ottawa, Canada
[2] Canada CIFAR AI Chair
Amii, Canada

The supplementary material is organized as follows: Section 1 gives some additional ablation studies, including the impact of different numbers of synthetic samples for the same background and the impact of key parameters. Section 2 provides additional qualitative evaluations to illustrate the effect of the proposed framework on the medical semantic segmentation task under the Exemplar Learning experimental scenario.

1 Additional Ablation Studies

1.1 Impact of Different Numbers of Synthetic Samples for the Same Background

We evaluate the performance of the number of synthetic samples for the same background in Table 1. Different numbers of synthetic samples are generated by ESM, which are subsequently used in different training stages to obtain the results. Clearly, an intermediate number, *i.e.*, 15, is beneficial for segmentation compared to $5, 10, 20, 25$. It is reasonable that, along with the increasing number of synthetic samples, the proposed framework is prone to overfitting around the space of the exemplar sample distribution. In this case, a certain amount of synthetic samples can offer enough guidance in our framework. Therefore, this setting can ensure that the number of synthetic samples considers sufficient and quality.

1.2 Impact of Key Parameters

We summarize the results by using the proposed framework with different values of the parameters of Eq. (8) on the ACDC and Synapse datasets in Figure 1. In this part, we set the parameter of main loss λ_e as one and fix others to obtain the result when adjusting a certain λ. As shown in Figure 1 (a), when the value of λ_s is set within a reasonable range, the experimental results can increase to some degree and reach their highest value at a setting of 1.2 (0.31 and 0.41). Then, when λ_c is set to 5 and 10, the best experimental results are achieved on the ACDC and Synapse datasets, respectively. Finally, better results are achieved when λ_u is set to smaller values while increasing it leads to a decrease in the overall trend. We analyse that this is mainly due to the degree of noise in the pseudo-labels.

Table 1: Ablation study of the different number of synthetic samples for the same background on the ACDC dataset. The results in different training stages are also presented.

Stage	Number	DSC.Avg↑	HD95.Avg↓	RV	Myo	LV
Stage ①	5	0.199	32.07	0.101	0.177	0.319
	10	0.300	34.38	0.186	0.279	0.435
	15	0.355	40.80	0.249	0.315	0.503
	20	0.323	29.55	0.207	0.303	0.460
	25	0.285	26.95	0.148	0.288	0.421
Stage ②	5	0.211	55.17	0.153	0.185	0.296
	10	0.320	22.02	0.138	0.333	0.488
	15	**0.410**	26.64	**0.293**	**0.374**	**0.563**
	20	0.377	**20.12**	0.279	0.360	0.491
	25	0.364	23.58	0.247	0.362	0.482

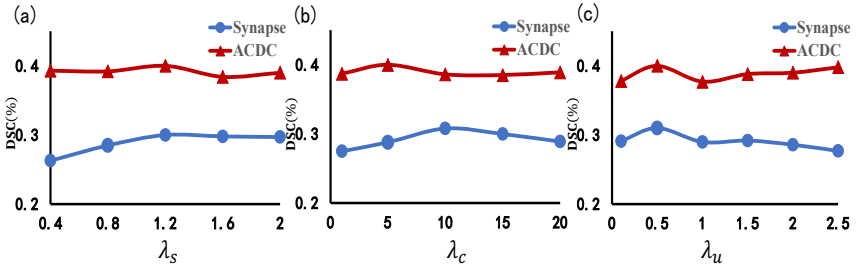

Figure 1: Analysis of hyperparameters on Synapse and ACDC datasets. In each of the three charts, the y-axis indicates the value of the DSC on the Synapse and ACDC datasets. The x-axis indicates the value of the parameters.

2 Additional Qualitative Results

2.1 Visualization of the Learned Embeddings

We present several visualized examples of learned embeddings in Figure 2 to demonstrate the effect of the proposed framework. The top row of each sample corresponds to the feature representing the feature maps focusing on the foreground, while the bottom row represents those focusing on the background. The selected features are from the output of the encoder. Red to blue in the figure indicates activation values from large to small. In the case of baseline, the features learned by the model are unable to focus on the target organ. When the proposed ESM module is considered, the feature map shows that the model is able to focus on the target region in a coarse manner. When the proposed ESM and PCEM modules are considered, the model can not only focus better on the target region, but its feature map focusing on the background can also clearly distinguish the background region.

2.2 Visualization Examples on Synapse and ACDC datasets.

We present additional qualitative results in Figure 3 and Figure 4. It can be seen that the proposed ELSNet is able to segment different organs in the same image. Note that we synthesise

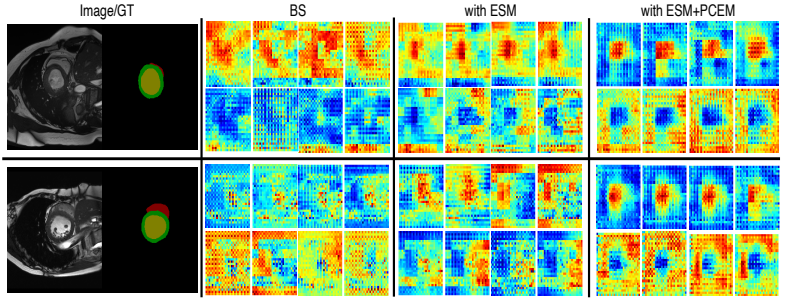

Figure 2: Visualization of the learned embeddings. "BS" denotes the results of the baseline. "with ESM" denotes the visualized embeddings obtained when only the proposed ESM module is used. "with ESM+PCEM" denotes visualized embeddings obtained when both of the proposed modules are used. Red to blue in the figure indicates activation values from large to small. The best view is in color.

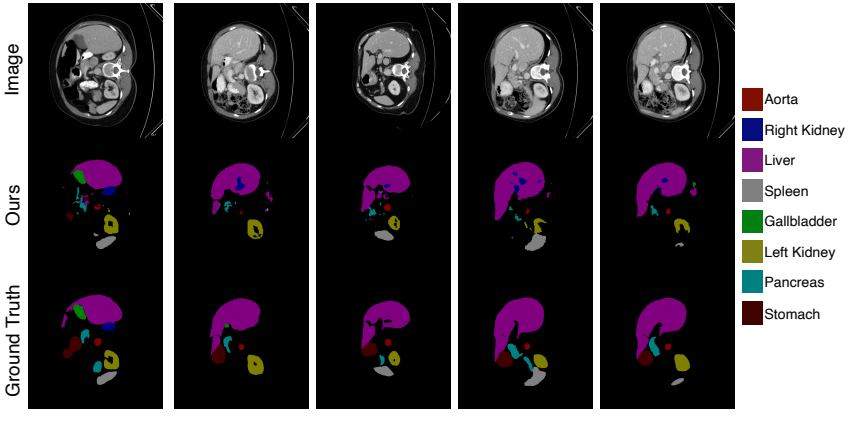

Figure 3: Visualized examples of our experimental results on the Synapse dataset. The first row indicates the images. The second row indicates the segmentation results by the ELSNet. The third row indicates the ground truth.

the dataset at a finer granularity so that the generated data is significantly different from the original data and simulates variation across the dataset, greatly enriching data diversity. Our proposed ESM generates synthetic data that are derived directly from real data and uses various transformations to mimic the appearance of the dataset. Meanwhile, our proposed ESM can obtain the ground truth for each pixel, overcoming the limitations of existing methods and making it particularly suitable for the segmentation task. Therefore, it is reasonable to learn from a single annotated exemplar to produce results.

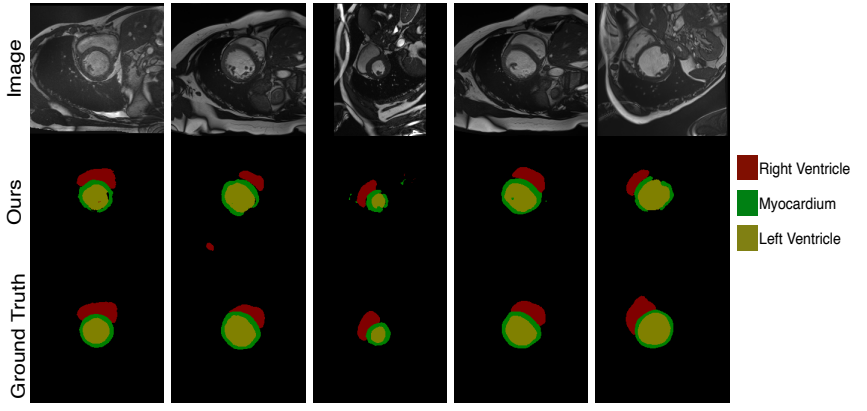

Figure 4: Visualized examples of our experimental results on the ACDC dataset. The first row indicates the images. The second row indicates the segmentation results by the ELSNet. The third row indicates the ground truth.