Self-supervision with Superpixels: Training Few-shot Medical Image Segmentation without Annotation

Cheng Ouyang[1(✉)], Carlo Biffi[1*], Chen Chen[1*], Turkay Kart[1*], Huaqi Qiu[1], and Daniel Rueckert[1]

BioMedIA Group, Department of Computing, Imperial College London, UK
c.ouyang@imperial.ac.uk

Supplemental Materials

Summary of the proposed algorithm

Algorithm 1 Training and testing of the proposed few-shot segmentation algorithm

---

Require: Training dataset $\{\mathbf{x}_i\}$, testing dataset $\mathcal{D}_{te}$, unsupervised superpixel algorithm $\mathcal{F}(\cdot)$ [68], feature extractor network $f_\theta(\cdot)$, Number of episodes N_{epi}, learning rate l_r.

1. Unsupervised pseudolabel generation:
for image $\mathbf{x}_i$ in $\{\mathbf{x}_i\}$ **do**
  Compute pseudolabel set $\mathcal{Y}_i = \mathcal{F}(\mathbf{x}_i)$.
end for
2. Training with pseudolabels:
for Episode i in range(N_{epi}) **do**
  Sample $\mathbf{x}_i \in \{\mathbf{x}_i\}$, $\mathbf{y}_i^r(c^p) \in \mathcal{Y}_i$ to form support $\mathcal{S}_i = \{(\mathbf{x}_i, \mathbf{y}_i^r(c^p))\}$.
  Apply random transforms to form query $\mathcal{Q}_i = \{(\mathcal{T}_g(\mathcal{T}_i(\mathbf{x}_i))), \mathcal{T}_g(\mathbf{y}_i^r(c^p)))\}$.
  Compute representation prototype ensemble $\mathcal{P}_i = \{p_k(c^j)\}$ from $\mathcal{S}_i$ using Equ. 1-3.
  Predict segmentation on $\mathcal{Q}_i$ with reference to $\mathcal{P}_i$ using Equ. 4-6.
  Compute loss $\mathcal{L}^i(\theta; \mathcal{S}_i, \mathcal{Q}_i)$ using Equ. 7-9.
  Update network parameters: $\theta^i = \theta^{i-1} - l_r \frac{\partial \mathcal{L}^i}{\partial \theta}$.
end for
3. Testing on real objects:
Obtain support $\mathcal{S} = \{(\mathbf{x}_l^s, \mathbf{y}_l^s(c^{\hat{j}}))\}$ and query $\mathcal{Q} = \{\mathbf{x}^q\}$ from $\mathcal{D}_{te}$.
Compute representation prototype ensemble $\mathcal{P} = \{p_k(c^j)\}$ from $\mathcal{S}$ using Equ. 1-3.
Predicte segmentation on $\mathcal{Q}$ with reference to $\mathcal{P}$ using Equ. 4-6.

---

*Equal contribution.

Generalizing learned image representations across domains

To evaluate the cross-domain generalization ability of the learned representations, we train the model on one dataset and evaluate few-shot segmentation on the other. Specifically, We experimented with Abd-MRI and Card-MRI: these two dataset are from two different MRI imaging sequences: the former is *T2-Spectral Presaturation with Inversion Recovery* (T2-SPIR) and the latter is *balanced steady-state free precession* (bSSFP). Also, they are sliced from different views (axial view for Abd-MRI and short-axis view for Card-MRI). The numbers reported below are Dice scores.

Table 1. Cross-domain few-shot segmentation performance

Trained on	Tested on	Mean Dice score
T2-SPIR	T2-SPIR	73.02
bSSFP	T2-SPIR	66.58 (-6.44)
bSSFP	bSSFP	76.90
T2-SPIR	bSSFP	60.52 (-16.38)

This result implies that image representations learned from SSL have reasonable generalization ability to a different domain.

Qualitative results

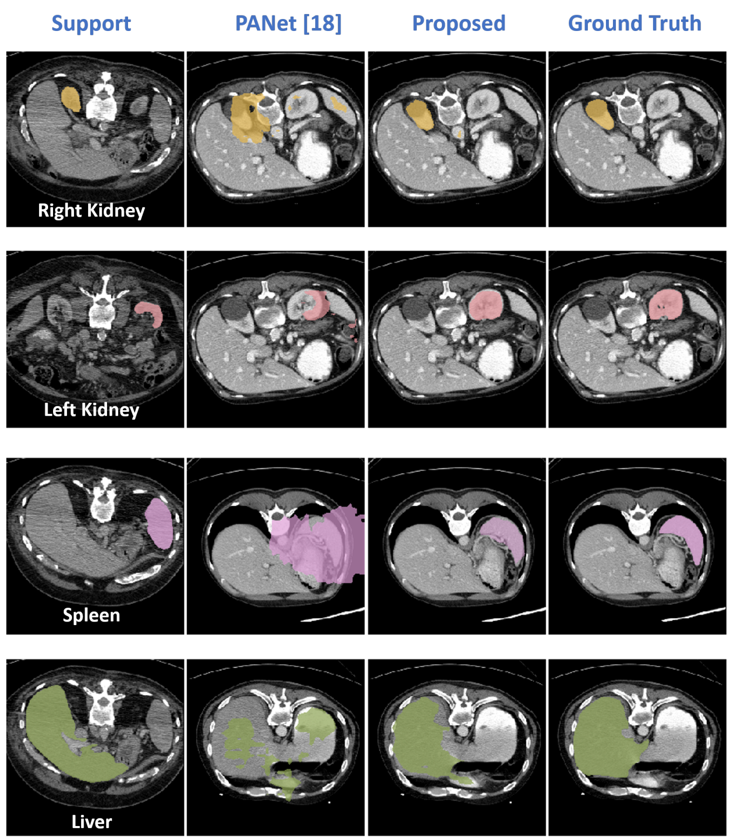

Fig. 1. Qualitative results of the proposed method on abdominal CT under setting 2, where images containing testing classes are strictly excluded in training set even though they are unlabeled.

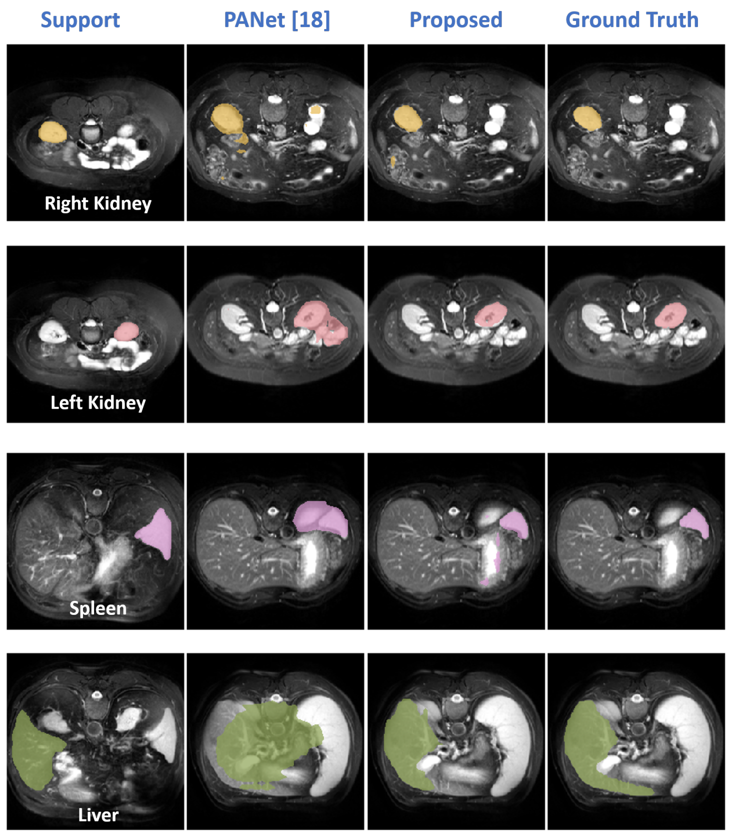

Fig. 2. Qualitative results of the proposed method on abdominal MRI under setting 2.

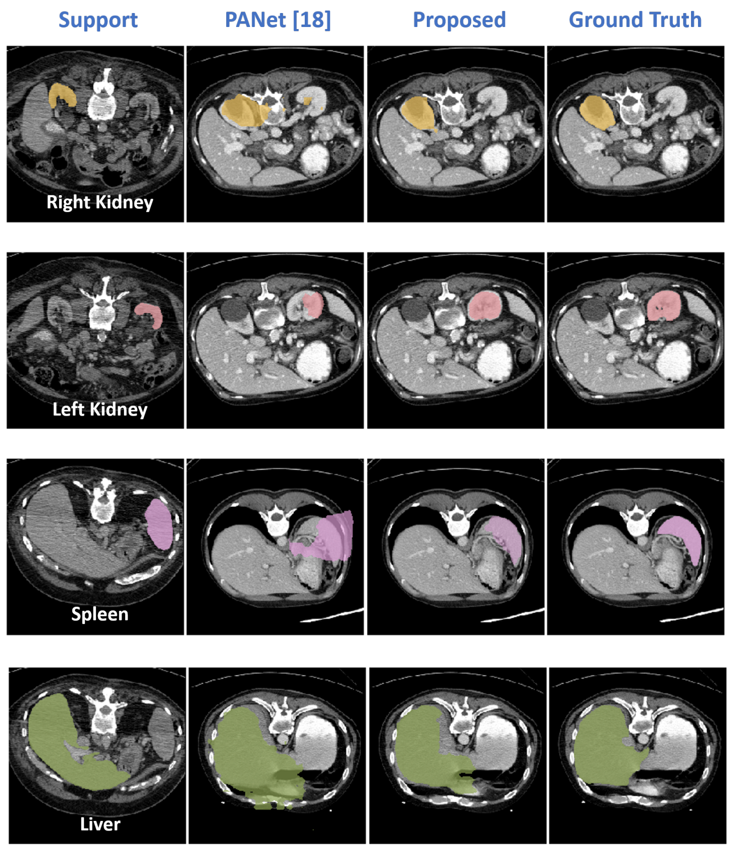

Fig. 3. Qualitative results of the proposed method on abdominal CT under setting 1, where objects of testing classes might appear in training images as part of the background.

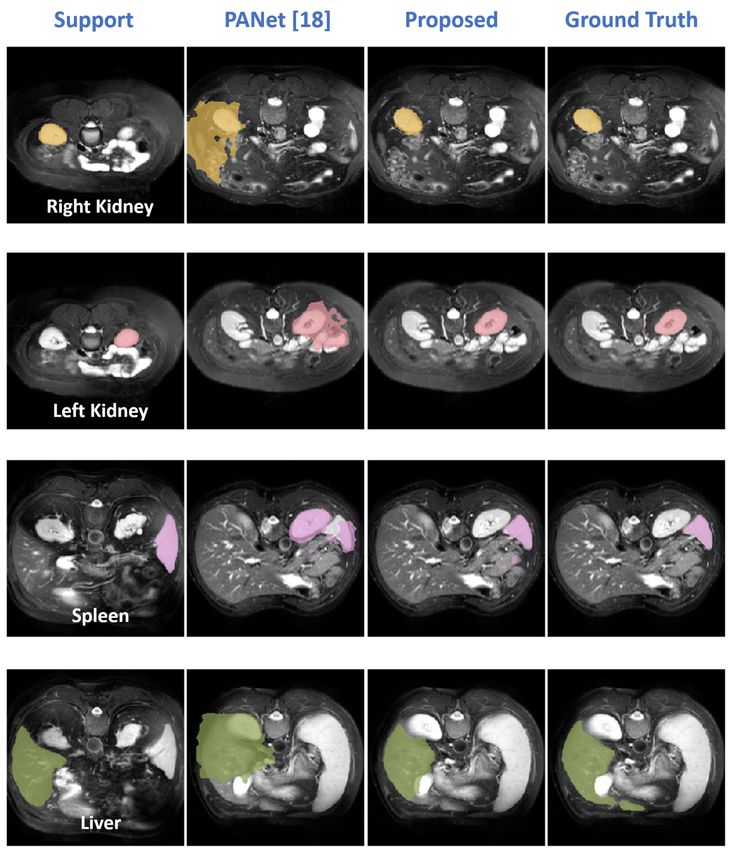

Fig. 4. Qualitative results of the proposed method on abdominal MRI under setting 1.

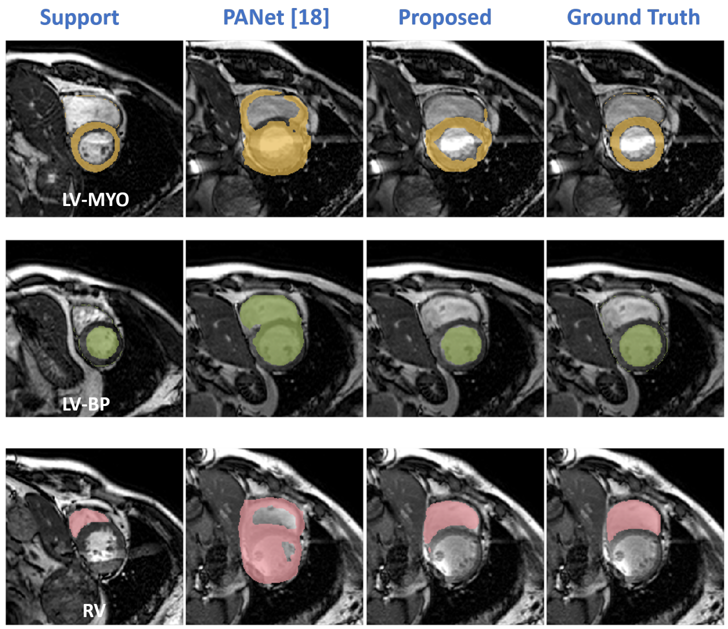

Fig. 5. Qualitative results of the proposed method on cardiac MRI under setting 1.

Examples of local and class-wise similarities

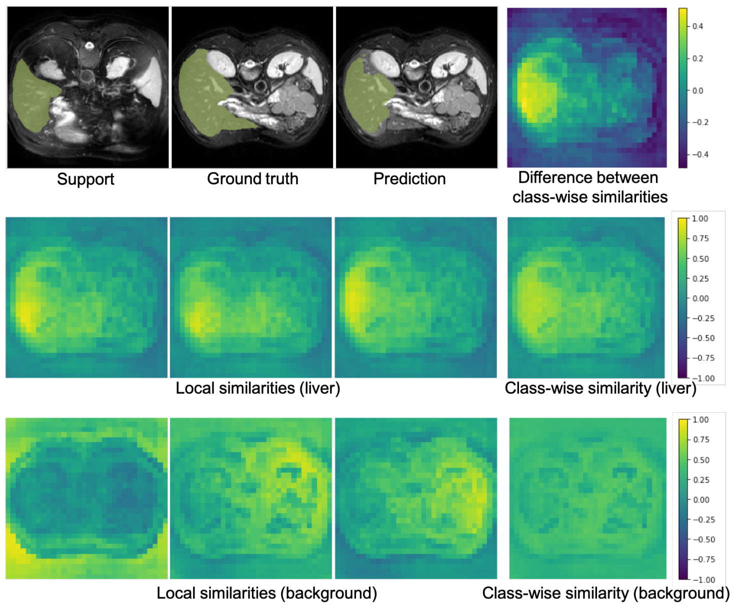

Fig. 6. Examples of local and class-wise similarities.

Local peaks in different local similarities differ with each other. This implies that one local prototype matches specific region(s) in the query. All these local similarities of a specific class are fused together to form a reasonable class-wise similarity map.

Illustration of the protocol for evaluating 2D FSS on 3D images

To evaluate 2D few-shot segmentation on 3D volumetric images, we follow the evaluation protocol established by [43]. In a 3D image, for each class, images between the top slice and the bottom slice containing this class are divided into C equally-spaced chunks. The middle slice in each chunk from the support scan is used for segmenting all the slices in the corresponding chunk in the query scan. In our experiments C is set to be 3.

For example, in Fig. 6, chunks are marked with blue (for query) or light purple (for support) boxes. Middle slices of the chunks the in support scan are marked in dark purple. Each of such middle slices serves as the support for all the slices in the corresponding chunk in query respectively. Of note, segmentation is performed on 2D **axial** slices, while in Fig. 6 the testing class *liver* is highlighted in purple on **coronal** plane for illustration purpose only.

As noted in [43], the top and bottom slices for each class can be found by a quick manual scroll-and-click process through 2D slices, or by using a simple model trained for this purpose.

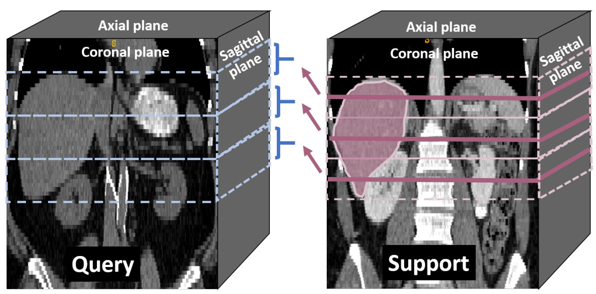

Fig. 7. Illustration of the protocol for evaluating 2D few-shot segmentation result on 3D volumetric images.