One-Shot Boundary Detection Network for Multi-Modal Side-Viewing Imaging

Beatriz Farola Barata[1,2,*], Guiqiu Liao[2,3,*], Diego Dall'Alba[2], Gianni Borghesan[1,4], Benoit Rosa[3], Michel de Mathelin[3], Florent Nageotte[3], Paolo Fiorini[2], Michalina J. Gora[3], Jos Vander Sloten[1], and Emmanuel Vander Poorten[1]

[1]*Robot-Assisted Surgery Group, Department of Mechanical Engineering, KU Leuven, Leuven 3001, Belgium*
[2]*Altair Robotics Laboratory, Department of Computer Science, University of Verona, Verona 37134, Italy*
[3]*ICube Laboratory, University of Strasbourg, Strasbourg, France*
[4]*Core Lab ROB, Flanders Make, Belgium*
[*]*These authors contributed equally to this work (Corresponding author: beatriz.barata@kuleuven.be)*

INTRODUCTION

Nowadays, catheter-based imaging systems are increasingly used in a variety of clinical applications in order to obtain side-viewing luminal and transmural images. Mainstream side-viewing catheters often use Intravascular Ultrasound (IVUS) or Optical Coherence Tomography (OCT) to acquire cross-sectional views of the intraluminal environment. IVUS is commonly used for imaging intravascular pathologies such as aneurysms or atherosclerotic plaque [1]. Side-viewing OCT can also be used to visualize vascular structures, as well as larger lumens, such as the colon or the respiratory lung airways. Both imaging modalities have their characteristics: IVUS can see through blood, but experiences problems seeing through air. OCT does not have that problem, but it relies on injection of saline solution or contact with tissue for imaging in blood. The resolution of OCT is typically higher than of IVUS whereas IVUS has a larger image depth. Despite their differences, in side-viewing catheters, both methods tend to produce images with a similar appearance. Automatic segmentation of both OCT and IVUS is an appealing feature for supporting real-time diagnosis or offline image analysis. However, the resemblance between these two imaging modalities has not yet been exploited to produce robust tissue segmentation algorithms. Traditional methods for segmentation of OCT or IVUS use handcrafted edge detection filters, mathematical morphology, Otsu's automatic thresholding, intersection of radial lines with lumen boundaries, Markov-random fields and light back-scattering [2].

This work proposes a novel deep Convolutional Neural Network (CNN) architecture, shown in Fig. 1, based on explicit coordinate encoding networks designed for multi-modal image processing. Aside from improved contour segmentation efficiency, the proposed approach directly provides the relative distance between tissue surface and scanning center. The proposed architecture directly encodes in a coordinates vector the position of the detected object contours in the polar domain, without pix-wise segmentation nets [3] or additional detection networks. Moreover, the same architecture is applied to OCT and IVUS images for the task of lumen segmentation and then compared with the state-of-the-art U-net architecture [3].

This work was supported by the ATLAS project. The ATLAS project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813782.

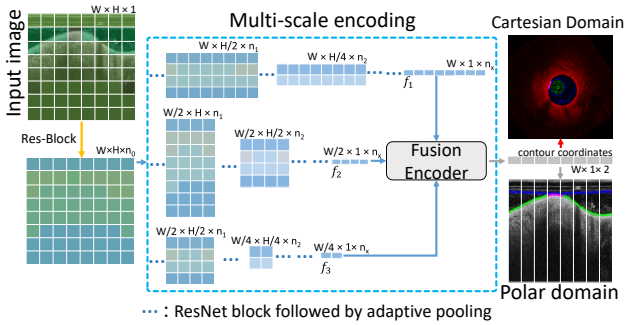

Fig. 1: Proposed network architecture featuring a parallel multi-scale encoding scheme.

MATERIALS AND METHODS

Inspired by shape encoding, the proposed network architecture predicts the contour coordinates of surgical objects and/or intraluminal structures. In particular, the final boundary coordinates prediction occurs in one shot and does not rely on detection nets or segmentation nets as backbone. The proposed method, shown in Fig. 1, first adapts an initial shallow Resnet block to produce raw features with the same dimension as the 2D input image (orange arrow in Fig. 1). Then, it follows a parallel multi-scale encoding scheme (blue dashed box in Fig. 1). Adaptive pooling modules, after identity Resnet blocks, extract multi-scale hierarchical coordinates position descriptors which potentially contain information for predicting different scales of contour coordinates. For instance, the first order of position descriptor $\boldsymbol{f}_1 \in \mathbb{R}^{W\times 1\times n_x}$ matches the width W of the input image $\boldsymbol{I} \in \mathbb{R}^{W\times H}$. $\boldsymbol{f}_1$ extracts features with higher spatial correspondence. In contrast, lower scale descriptors $\boldsymbol{f}_2$ and $\boldsymbol{f}_3$ represent the position information with lower scale in the horizontal direction, but they extract more abstract features that are less sensitive to noise. Note that Fig. 1 just illustrates the schematic of the proposed networks. Applications can thus have more than 3 coordinates descriptors $\boldsymbol{f}_i$ (i=1,2,3...). The higher parallelism allows for much faster inference compared to extracting hierarchical information in a cascaded way, as implemented in the U-Net architecture. The fusion encoder (gray block in Fig. 1) is deployed after the sub-branches to re-organize the multi-scale information,

TABLE I: Comparison with post-processed U-net on the different datasets with both region- and boundary- based metrics

	Jaccard index(↑)			Dice coefficient (↑)			Boundary error/pixel (↓)	
	Lumen	Catheter	Tissue	Lumen	Catheter	Tissue	R1 v R2	R2 v R3
Phantom IVUS UNet	0.953±0.032	0.970±0.014	0.984±0.013	0.976±0.018	0.985±0.007	0.992±0.007	2.32±0.34	3.39±2.36
Phantom IVUS Ours	**0.974±0.017**	**0.988±0.006**	**0.992±0.006**	**0.987±0.009**	**0.994±0.003**	0.996±0.003	**0.26±0.11**	**1.42±1.04**
Phantom OCT UNet	0.973±0.035	0.937±0.043	0.600±0.297	0.985±0.019	0.966±0.025	0.699±0.242	2.54±0.38	5.51±11.2
Phantom OCT Ours	0.988±0.013	**0.984±0.011**	**0.851±0.105**	0.993±0.007	**0.991±0.005**	**0.915±0.065**	**0.73±0.40**	**1.87±1.59**
In vivo OCT UNet	0.704±0.117	0.916±0.019	0.894±0.051	0.820±0.092	0.956±0.010	0.943±0.029	2.06±0.47	11.8±9.84
In vivo OCT Ours	**0.918±0.088**	**0.946±0.015**	**0.958±0.070**	**0.954±0.066**	**0.972±0.007**	**0.976±0.059**	**0.99±0.28**	**4.39±2.83**

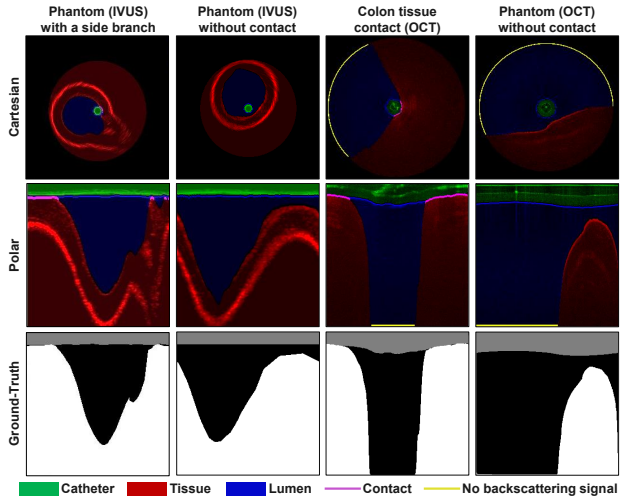

Fig. 2: Qualitative segmentation results.

which aligns all lower scale features to the original image scale, and produces A-line level contour coordinates as the boundary detection result.

RESULTS

An IVUS probe embedded at the tip of a robotic catheter with an active distal segment was steered in a poly(vinyl alcohol) (PVA) cryogel vessel phantom to collect the first dataset (3500 images). OCT images were acquired by steering an OCT probe in a colon phantom with layered tissue [4] (3000 images) as well as in an *in vivo* swine colon [5] (2000 images).

Qualitative results: Fig.2 shows representative results of the boundary detection experiments. Segmentation results are shown both in Cartesian and Polar domain. The detected boundaries are used to segment images. The catheter is shown in green, the tissue area is marked red, lumen (air or blood) is shown in blue and contact between catheter/sheath and tissue is highlighted in purple. Finally, A-lines without back scattering are marked yellow. The first column of Fig. 2 shows the IVUS catheter in contact with the tissue and close to a side branch, visible in the image. The second column shows a case where the catheter with IVUS is inside the phantom (small lumen) with good lumen contrast and no contact with the tissue. The third column is taken with OCT in a large lumen. The catheter is pressed in a tight contact between plastic sheath and tissue. The fourth column, in the same large lumen, shows the tissue mostly as a flat surface with the rest of the image showing background noise.

Validation strategies: For both IVUS and OCT, the acquired images were split into train and test dataset by 2:1. The region segmentation accuracy was computed by means of the Jaccard index and the Dice coefficient, comparing the proposed method with a U-net trained with GAN loss [3]. As shown in Table I, the proposed architecture shows higher accuracy for all datasets. Furthermore, the U-net does not directly output the contour position. Through post-processing of the boundaries extracted from the U-net, the Euclidean distance with the ground-truth boundary was computed as an additional validation metric. Errors for 2 boundaries between 3 regions were calculated (regions of catheter, tissue and lumen). The last two columns of the Table I show the pix errors of each boundary ("R1 v R2" and "R2 v R3" denote boundaries of catheter/lumen and lumen/tissue, respectively). The proposed method has significantly lower errors compared to the prediction based on U-net. Mean distance errors are reduced from 5.4±5.5 to 1.9±1.2 for OCT images, and from 2.9±1.8 to 0.8±0.9 for IVUS images.

CONCLUSIONS AND DISCUSSION

In this work, a novel boundary detection architecture is proposed for polar domain processing of images acquired from side-viewing catheters. The proposed approach is applied to two different imaging modalities and evaluated by both area and boundary based metrics. The results show that, in comparison to state-of-the-art methods, higher accuracies are obtained for both OCT and IVUS. Moreover, no post-processing is required to predict the contour coordinates. The proposed method showed superior performance with accuracy improvements by 20% for side-viewing catheter images. Future work will investigate the use of this architecture for other tasks requiring boundary prediction such as e.g. for inspection of the esophagus by a capsule catheter. Further improvements could follow from predicting the existence probability for every object per A-line.

REFERENCES

[1] Chaoyang Shi *et al.*, "Three-dimensional intravascular reconstruction techniques based on intravascular ultrasound: A technical review," *IEEE journal of biomedical and health informatics*, vol. 22, no. 3, pp. 806–817, 2018.

[2] Yan Ling Yong *et al.*, "Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography," *Journal of biomedical optics*, vol. 22, no. 12, p. 126005, 2017.

[3] Isola Phillip *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[4] Natalia Zulina *et al.*, "Colon phantoms with cancer lesions for endoscopic characterization with optical coherence tomography," *Biomedical optics express*, vol. 12, no. 2, pp. 955–968, 2021.

[5] Oscar Caravaca Mora *et al.*, "Steerable oct catheter for real-time assistance during teleoperated endoscopic treatment of colorectal cancer," *Biomedical optics express*, vol. 11, no. 3, pp. 1231–1243, 2020.