Segmentation of Lumen in Ureteroscopy Images using Deep Learning

Jorge F. Lazo[1,2], Aldo Marzullo[3], Sara Moccia[4,5], Benoit Rosa[2], and Elena De Momi[1]
[1] *Department of Electronics,* [2]*Information and Bioengineering*, Politecnico di Milano, Milan, Italy
[2] *ICube*, Université de Strasbourg, Strasbourg, France‘
[3] *Department of Mathematics and Computer Science, University of Calabria, Cosenza, Italy*
[4] *Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy*
[5] *Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy*
jorgefrancisco.lazo@polimi.it

I. INTRODUCTION

Ureteroscopy is a minimally invasive procedure which is used to explore the upper urinary tract, allowing the diagnosis and treatment of different conditions, such as kidney stones or urothelial carcinoma. Navigation and diagnosis inside the urinary tract are highly dependent upon the operators experience, and some image-related conditions such as the presence of image artifacts, floating debris, occlusions in the video, or image noise could add additional challenges for non-experienced operators. The development of computer vision methods for surgical assistance aims to deal with these limitations by highlighting relevant information which can enhance the performance of the surgeon and minimize the probability of complications.

The task of lumen segmentation is a fundamental part in the development of these assistance systems since this is the reference which marks the path that the endoscope should follow. However, this is not a simple task since there is a high variability in the the inter-patient anatomical structures, as well as the inner-patient variability, in terms of the different shapes in which the lumen is deformed at different points in the urinary tract.

Previous implementations of lumen segmentation has been proposed for colonoscopy images, being the majority of them based on the design of handcrafted feature extractors which require the tuning of several parameters [1], [2]. The most recent study proposes the use of a FCN-8 network [3] for the general segmentation of elements appearing of colonoscopy images, among them the lumen.

To the best of our knowledge, Convolutional Neural Networks (CNNs) have not been applied to semantic segmentation of the lumen in ureteroscopy images. We attribute this with the lack of a publicly available annotated dataset of ureteroscopy images, which is needed in order to train and validate such kind of networks. To tackle this we collected and annotated our own dataset from ureteroscopy images.

In this paper we propose the application of a variation of the well-known CNN architecture U-Net for the task of lumen segmentation by adding batch normalization layers at the output of each of the convolutional layers and changing the loss function to the Dice similarity coefficient loss (L_{DSC}).

II. PROPOSED METHOD

Inspired by the model proposed in [4] originally developed for axon segmentation in microscopy images, we propose the implementation of a lumen-segmentation network based on U-Net architecture. We applied batch normalization at the output of every convolutional layer. In our specific case, the use of batch normalization provided stability and convergence of the learning process given the high variability in the structures, brightnes and image resolution among the images of the dataset we used. A set of experiments was carried out using the original implementation of U-Net without batch normalization. It was observed that during training without batch normalization, without a proper initialization of the weights the network got stuck in a point where only white, or only black images were obtained as outputs and with DSC values bellow 0.25 in average.

The second modification with respect to [4] is a change in the chosen loss function. The one used in this implementation was the (L_{DSC}) defined as:

$$L_{DSC} = 1 - \frac{2TP}{2TP + FN + FP} \tag{1}$$

where TP is the number of pixels that belong to the lumen, which are correctly segmented, FP is the number of pixels miss-classified as lumen, and FN is the number of pixels which are classified as part of lumen but actually they are not. Adam optimization was used during the training. The learning rate, and mini batch size for each of the models was chosen by trying the different combinations between several possible values of the hyper-parameters and using a 5-fold cross validation strategy. Once the hyper-parameter values were chosen, the training process was done dividing the data set for training/validation in a ratio 65/35. The Wilcoxon T-test on the DSC was used to determine statistical significance between the different models trained.

For this study, 7 video dataset from 4 patients were collected. The videos were acquired from the European Institute

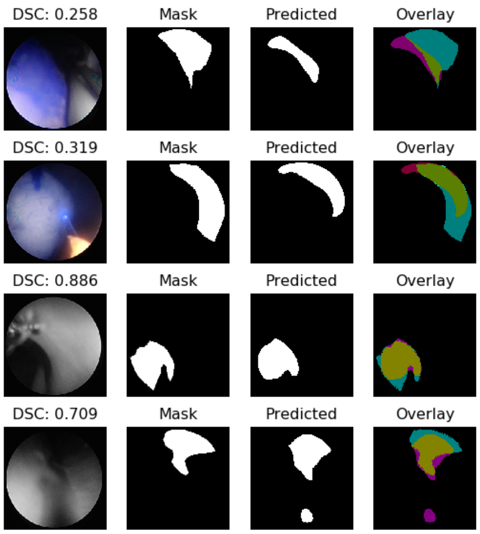

Fig. 1: Sample of results obtained using U-Net with BN with its respective DSC score. The colors in the Overlay images are as follows, TP: Green, FP: purple, FN: Blue, TN: Black.

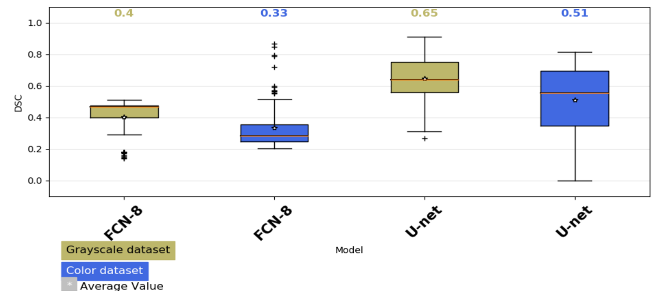

Fig. 2: Box plots of DSC obtained with the different CNNs tested.

of Oncology (IEO) at Milan, Italy and the KU Leuven University Hospital. All patients gave their informed consent for data collection and use of it for research. The data collection followed the ethical protocol approved by the IEO and in accordance with the Helsinky Declaration. The videos are from ureteroscopy procedures targeting upper tract tumor ablation and kidney stone removal. From these videos, a total number of 1,445 frames were extracted and manually annotated using our own GUI developed in Matlab for this purpose. The video-frames from patient one were set apart to be used as test dataset. We compare the obtained results with the FCN-8 architecture that has been used for lumen segmentation in colonoscopy images [5].

The performance metrics chosen to quantitatively evaluate the results, were the DSC, the Precision ($Prec$) and Recall (Rec), which are defined as

$$DSC = 1 - L_{DSC} \tag{2}$$

$$Prec = \frac{TP}{TP + FP} \tag{3}$$

$$Rec = \frac{TP}{TP + FN} \tag{4}$$

III. RESULTS AND DISCUSSION

The boxplots of the DSC are depicted in Fig. 2. From these result is possible to see that U-Net has the best performance overall. The average values of DSC, $Prec$ and Rec for this network were of 0.65, 0.48 and 0.82 respectively. In comparison with FCN-8 in the gray-scale dataset U-Net achieves a DSC 0.25 better than the FCN-8 network ($p<0.05$) and in the case of the RGB dataset achieves an average value 0.18 ($p<0.01$) better than FCN-8. In general for both models it was seen that the training in gray-scale images achieve better results among the same models. In the case of U-Net a difference of 14% was observed ($p<0.001$) while in the case of FCN-8 the difference was of 7% ($p<0.05$). This might be related to the color-space itself of the dataset. Considering that the gray-scale image dataset might have enough information to perform the binary classification of pixels into lumen and no-lumen, the addition of two channels would require the adjustment of more parameters, which implicitly would require a higher amount of data. Some sample of the segmented images are depicted in Fig. 1

IV. CONCLUSION

In this paper, we addressed the task of lumen segmentation in the ureter, for this purpose we proposed the implementation of a Deep CNN. The network used consisted of a standard version of U-Net with the addition of batch normalization at the output of each of the convolutional layers. The method was tested in different image spaces finding that for this task the network performs better in gray-scale images and then compare with a FCN-8 model which previously has been used for lumen segmentation in colonoscopy images. We show that the network performs better than the previous model implemented for this task. This demonstrates that it could be a suitable model for further development in the task of lumen segmentation of the ureter.

ACKNOWLEDGEMENT

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

REFERENCES

[1] G. Gallo and A. Torrisi, "Lumen detection in endoscopic images: a boosting classification approach," *International Journal On Advances in Intelligent Systems*, vol. 5, no. 1, 2012.

[2] D. Wang, X. Xie, G. Li, Z. Yin, and Z. Wang, "A lumen detection-based intestinal direction vector acquisition method for wireless endoscopy systems," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 3, pp. 807–819, 2014.

[3] J. Verma, M. Nath, P. Tripathi, and K. Saini, "Analysis and identification of kidney stone using k th nearest neighbour (knn) and support vector machine (svm) classification techniques," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 574–580, 2017.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[5] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdzal, and A. Courville, "A benchmark for endoluminal scene segmentation of colonoscopy images," *Journal of Healthcare Engineering*, vol. 2017, 2017.