Comparison of Deep Learning Architectures for Lumen Segmentation

Jorge F. Lazo, Aldo Marzullo, Sara Moccia, Benoit Rosa & Elena De Momi & Michel de Mathelin

Université de Strasbourg

Politecnico Milano

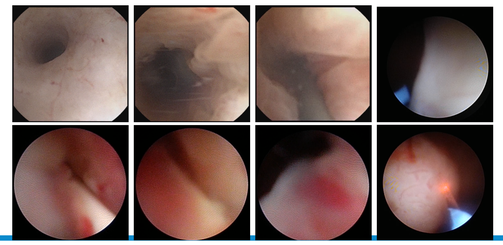
Abstract

The task of lumen segmentation is a fundamental part in the development of computer vision methods for surgical assistance, since this is the reference which marks the path that the surgeons should follow with the endoscope in minimal invasive procedures. We study the implementation of different models of Deep Neural Network for the task of lumen segmentation. For the training of these networks, we analyze the use of two different color spaces: gray-scale and RGB data images and also converting the images to polar coordinates.

Keywords: deep learning, ureteroscopy, convolutional neural networks, lumen segmentation.

Introduction

Ureteroscopy is a procedure dedicated to the exploration of the upper urinary tract to perform diagnosis and treatment of different conditions, such as kidney stones or carcinoma of the upper urinary tract.
Navigation and diagnosis inside the urinary tract are highly dependent upon the operators expertise, and image-related conditions such as the presence of image artifacts, floating debris, the low quality of the video,etc. could add additional challenges for non-experienced operators.



Materials & Methods

For this study 7 videos were collected and from them selected frames were manually annotated. The videos were acquired from the European Institute of Oncology (IEO) at Milan, Italy. The loss function used in this implementation was based on the dice similarity coefficient (L_{DSC}), and the DSC as the performance metric to compare the models. The Kruskal-Wallis test on the DSC was used to determine statistical significance between the different models trained.

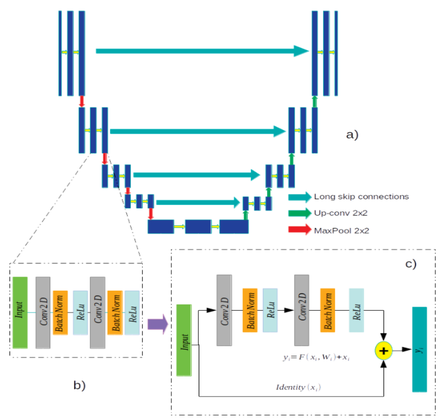
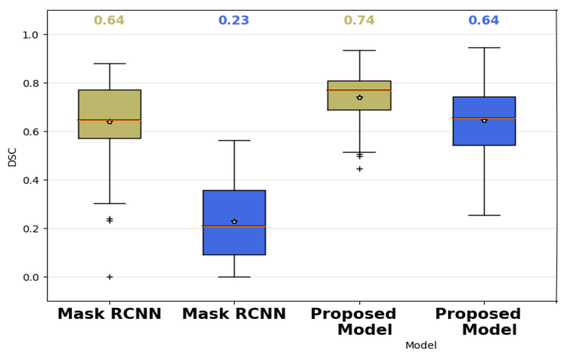


Figure 3: Proposed model

Results 1

In a second stage the model was compared with the Mask-RCNN architecture. Additionally the images went through a polar transformation before being feed to the networks.

- In general it was observed that the proposed model performs better than Mask-RCNN
- In general the transformation of the frames to polar coordinates did not improved the performance of any of the networks tested.



Results

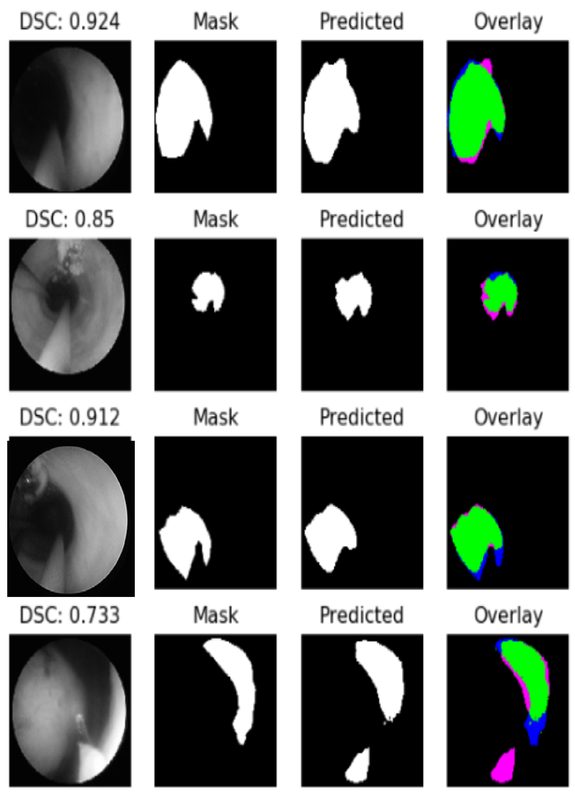


Figure 1: Sample of Results obtained with the proposed model based on residual blocks

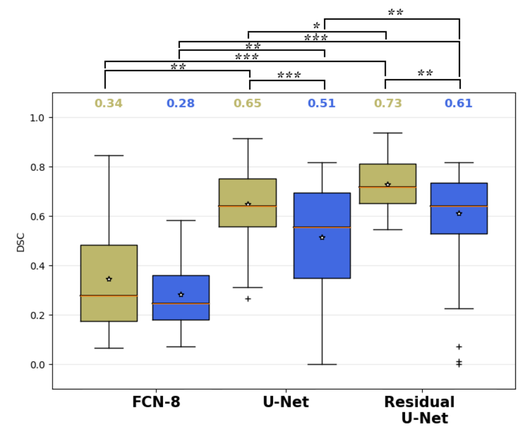


Figure 2: Box plots of the DSC score in the dataset for the three models tested initially

The box plots of the DSC, $Prec$ and Rec are depicted in Fig. 2 for each of the networks tested. From these results is possible to see that Residual U-Net has the best performance overall in DSC, $Prec$ and Rec with average values of 0.73, 0.58, and 0.92 respectively. In general for all the 3 models it was seen that the training in gray-scale images achieve better results among the same models. In the case of the simple U-Net a difference of 14% was observed ($p<0.001$) while in the case of Residual U-Net the difference was of 12% ($p<0.05$).
Further development to properly segment the complete area of the lumen is needed and this could be achieved by implementing more complex models

Conclusion

- Residual U-Net has the best performance overall in DSC, $Prec$ and Rec with average values of 0.73 0.73, 0.58, and 0.92 in comparison with the FCN-8 and the standard version of U-Net.
- Changing the image data to polar coordinates in general did not showed improvements in the segmentation obtaining lower scores
- The use of the Mask-RCNN in general was more complicated, so further research needs to be carried in order to find the optimal hyperparameters before discarding the use of this architecture for lumen segmentation.

References

[1] B. et al Chew. Robotic ureteroscopy: The future of stone management? *Urology Times*, 2020.

[2] S et. al Ali. An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. *Scientific reports*, 10, 2020.

Future Research

Future work includes the exploitation of temporal features embedded in the collected videos along with the collection of more data. The exploration of other architectures to better tackle the segmentation of the lumen and specifically the border regions in which is hard to define where the lumen begins. There is also the possibility of adding extra annotated frames so temporal features could be exploited.

Contact Information

Email jorgefrancisco.lazo@polimi.it
Phone +39 334 2427 146