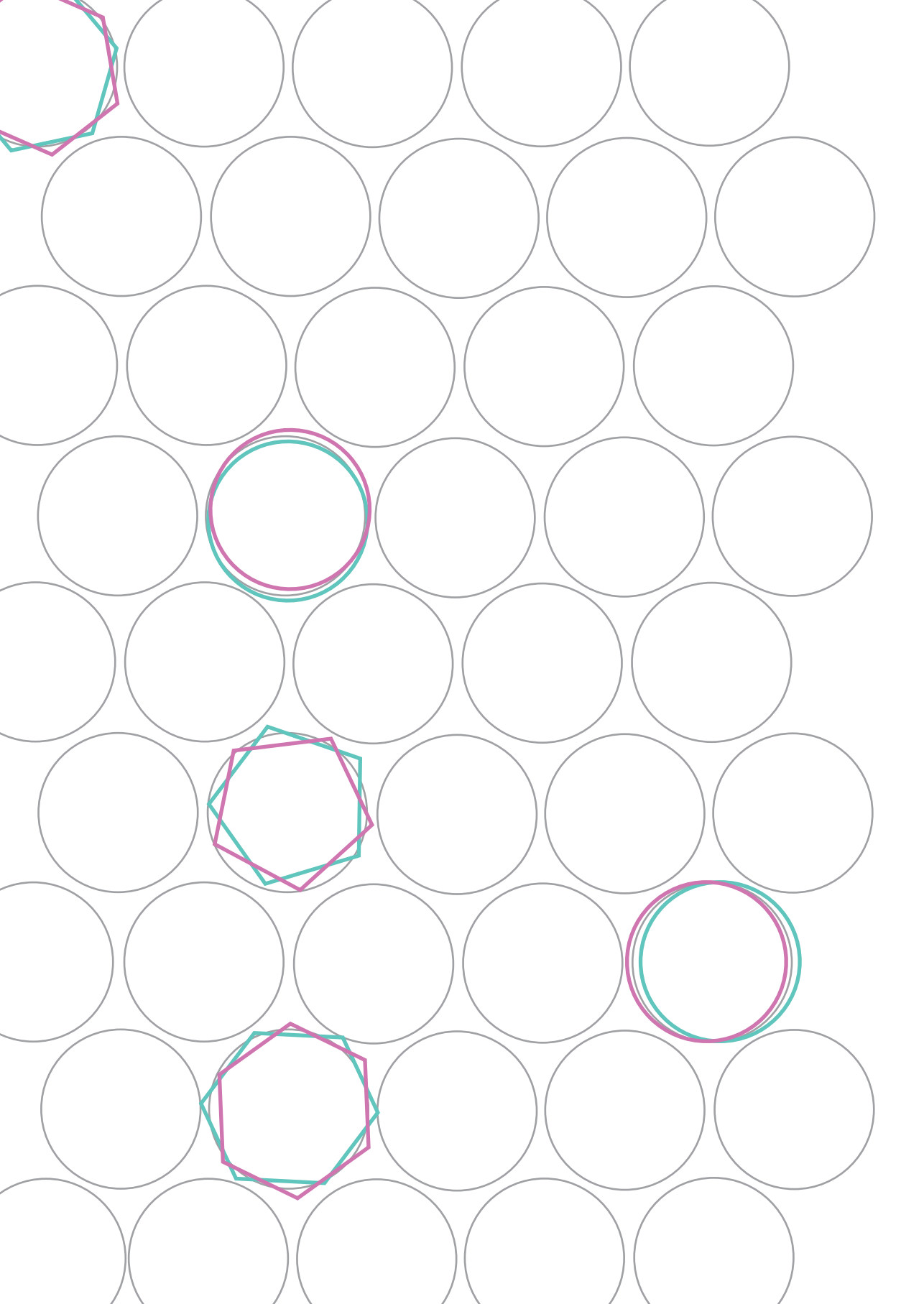
Deep learning for automatic segmentation of tumors on MRI

Rodríguez Outeiral, R.



Appendices

SUMMARY

Tumor segmentation is a crucial part of the radiotherapy treatment workflow. In the current clinical practice, this task is done manually by expert clinicians. This is time-consuming. A promising alternative is to automate this task with deep learning (DL) techniques. These methods have already been successfully applied clinically to segment other relevant structures for the radiotherapy treatment planning, such as the organs at risk. However, the automatic segmentation of tumors is not yet part of the clinical workflow. The aim of this thesis was to implement DL techniques to deliver clinically acceptable tumor segmentations. These algorithms were applied in two different MRI-based cohorts: a cohort of oropharyngeal primary tumors in multiparametric diagnostic MRIs and a cohort of cervical cancer gross tumor volume in MRI images of brachytherapy treatment.

When clinicians delineate the oropharyngeal cancer, they often rely on the information from several MRI sequences, such as the T1 weighted after gadolinium injection, T2 weighted or T1 weighted sequences. We hypothesised that DL methods may also benefit from combining the information from these different MRI sequences. Therefore, in **Chapter 2**, we studied the effect of combining different MRI sequences as input for the segmentation networks. Indeed, the performance of the DL networks improved when combining different MRI sequences as inputs compared to the use of single sequences.

Given the complexity of the task of oropharyngeal cancer segmentation, we hypothesised that we could achieve better auto-segmentations by simplifying the task. To that end, we split the task in two stages: first a coarse localization of the tumor and then its fine segmentation. In **Chapter 2**, we implemented this two stage approach in a semi-automatic manner. The first stage consisted of a localization step manually performed by clinicians, who were asked to draw a box around the tumor. Then, the final segmentation was performed by a DL network within the drawn box. As expected, the auto-segmentations rendered by this semi-automatic two stage approach outperformed the auto-segmentations achieved when directly providing the whole image as input.

The oropharyngeal cancer is present in a substantially smaller amount of voxels compared to the rest of structures in the MRI image. In the field of DL for automatic segmentation, this is known as the class imbalance problem, and can result in poor segmentation performance. In **Chapter 3**, we studied two different strategies to tackle this problem. One of these strategies was the use of different loss functions to train the segmentation networks. The most commonly used loss function when training segmentation networks is the Dice loss function, even though it is known to be suboptimal for smaller structures. In this chapter, we trained the segmentation networks with different loss functions designed to tackle class imbalance. Our results showed that all the proposed loss functions performed comparably to each other for the case of oropharyngeal cancer segmentation.

Another strategy to tackle the class imbalance problem consisted of the implementation of a two stage approach. Similarly to the previous chapter, we split the task in a localization step and a segmentation step. However, in **Chapter 3**, both stages were performed by neural networks thereby fully automating the semi-automatic approach proposed in the previous chapter. We demonstrated that our proposed two stage approach was an effective strategy to mitigate the class imbalance problem.

Brachytherapy is part of the standard of care for locally advanced cervical cancer. In this type of therapy, an applicator is placed inside the patient for treatment delivery. The tumor segmentations are made with this applicator inside, which can be uncomfortable for the patient. Therefore, it is desirable to acquire the tumor segmentations as promptly as possible. Due to these time constraints, the need for automatic segmentation is even more critical in the case of brachytherapy. In **Chapter 4**, we assessed the quality of the automatic segmentations of the cervical cancer gross tumor volume on brachytherapy MRI images. This assessment was performed both geometrically and dosimetrically. Our results showed similar dose-volume parameters as the manual segmentations used clinically, indicating that current DL methods can already render close to clinically valid tumor auto-segmentation in some cases.

In other cases, DL methods still produce tumor auto-segmentations that would not be clinically valid. Consequently, clinicians would still need to verify whether these auto-segmentations are acceptable for clinical use, limiting the time-gains of automatic segmentation methods. Therefore, there is a need for metrics that describe the quality of the auto-segmentations. However, common metrics to assess the quality of auto-segmentations also rely on comparing them to manually drawn segmentations, making them unsuitable for quality assurance. In **Chapter 5**, we identified a quality metric that can be generated directly from the output of the network. This quality metric had a high capability to distinguish between well and poorly performing auto-segmentations, showcasing its potential for quality assurance.

In conclusion, in this thesis we implemented different DL based approaches for automatic segmentation of tumors on MRI images. Strategies that provided relevant prior information to the segmentation networks were proved effective to increase the quality of the auto-segmentations, such as combining different MRI sequences as input or restricting the context around the tumor. Furthermore, we illustrated that current auto-segmentation frameworks can already render auto-segmentations that are comparable to clinically valid segmentations. Finally, besides the demonstrated improvements in the quality of the tumor auto-segmentations, we proposed a quality assurance metric that can distinguish between well and poorly performing cases. This type of metrics will potentially play a crucial role in advancing auto-segmentation methods for tumors towards clinical applicability.