Automatic Morphology-Based Detection of Hip Cartilage Segmentation Failure

Rebecca Poonam Brachat

Supervisors: MSc Adrian Ruckli, PD Dr. Kate Gerber
Institution: University of Bern , School of Biomedical and Precision Engineering
Examiners: MSc Adrian Ruckli, PD Dr. Kate Gerber

Introduction

Femoroacetabular impingement (FAI) and developmental hip dysplasia (DDH) are the leading cause of cartilage erosion resulting in severe osteoarthritis (OA) if left untreated [1]. Cartilage quality is an important diagnostic factor in treatment planning and a major predictor of surgical outcome. Thus, it can be used to identify patients who would benefit more from minimally invasive correction or from total hip replacement. With automatic 3D modelling, the cartilage composition, surface area, and volume can be calculated. However, automatic segmentation can fail and therefore in a clinical setting requires manual review by the clinician. Currently, segmentation failure is detected by either generating an uncertainty map or predicting a segmentation quality metric, like the Dice Similarity Coefficient (DSC). However, these methods would fail in cases where the uncertainty is low, or the DSC is high, but the morphology of the segmentation is suspicious (see Fig. 1), requiring manual review by a clinician. Therefore, morphology-based detection of segmentation failure would reduce manual review time and increase confidence in the accuracy of the automatic segmentation; however, no such system has been reported so far.

Materials and Methods

A classification network was trained and validated with N=243 cartilage segmentations of diagnostic MP2RAGE dGEMRIC images of 60 symptomatic FAI and DDH patients. First, a shape-correction model was trained to detect and correct suspicious morphology. With transfer learning, two different classifiers were trained to classify segmentations into "accurate" and "inaccurate" based on the learned features from the shape-correction model. The best performing classifier was tested with a test set comprised of the automatic segmentations of MP2RAGE, 1.5T, and 3T Dual-Flip dGEMRIC images. The Dual-Flip images were out of distribution for the used segmentation network, thus forcing the generation of segmentations of poor quality. The outcome measures for the shape-correction model is the DSC and for the classifier accuracy, sensitivity, and specificity.

Results

The shape-correction model achieved a DSC of 0.83 ± 0.04. The shape-correction model is partially able to detect and correct erroneous areas of the segmentation.

Fig. 1: Left: Suspicious automatic segmentation with protrusion on the right. Centre and right: Result of shape-correction model; ground truth (centre) and overlay (orange) of suspicious input image and prediction (yellow).

The classification network with the best performing classifier reached an accuracy of 49.4%, a sensitivity of 50.0%, and a specificity of 48.0%. The classification network was therefore able to detect approx. 50% of all suspicious cartilage segmentations.

Discussion

In this thesis, the first fully automatic morphology-based segmentation failure detection system is presented. The proposed classification network does not reach satisfactory accuracy, sensitivity, and specificity to be employed in clinical routine. Limitations are the generation of inaccurate segmentations for training and avoiding bias in the shape-correction and classifier model to a certain erroneous feature. Furthermore, the question of what is considered accurate and inaccurate cartilage morphology remains. Although the network requires further improvement, the idea of detecting segmentation failures based on morphology remains a promising idea.

References

N. C. Casartelli et al., “Is hip morphology a risk factor for developing hip osteoarthritis? a systematic review with meta-analysis,” Osteoarthritis and Cartilage, vol. 29, no. 9, pp. 1252–1264, 2021.

Acknowledgements

The project was supported by Florian Schmaranzer and Malin Meier, University hospital of Bern (Inselspital). They provided the hip data and the respective manual segmentations.