Quest for a Clinically Relevant Medical Image Segmentation Metric: the Definition and Implementation of Medical Similarity Index.

Szuzina Fazekas, Viktor Bérczi, Pál Maurovich Horvat, Zsolt Vizi
Semmelweis University, Budapest, Hungary

Introduction: In the field of radiology, an emerging number of images are created every day. In the medical field, delineation of different tissues and organs has a crucial role in diagnostics and therapeutics. The gold standard is manual segmentation by an expert, but nowadays there are more and more machine learning-based automatic segmentation methods. Thus, there is an understandable need for the quantification of the accuracy of a current segmentation. There are different widely used area-based and distance-based metrics, which are used for the evaluation of the accuracy of segmentations. These metrics only incorporate geometrical properties and fail to adapt to different clinical applications.

Aims: Our aim was to define and implement a clinically relevant medical image segmentation metric which has the opportunity to adapt to different clinical applications.

Methods: The reference contour was consider as the gold standard segmentation, and the agreement of a test contour to the reference contour was quantified. The bidirectional local distance was defined, and based on this distance, the points of the test contour were paired to points of the reference contour. After correcting with the distance between the test and reference center of mass, the euclidean distance was calculated between the paired points and a score was given to each of the test points. The overall medical similarity index was calculated as the average of the scores along all the test points.

Result: We created an image processing pipeline in Python programming language. The code is available at a public GitHub repository, and we also provide a Google Colaboratory notebook. The algorithm can handle multi-slice images with more than one mask in one slice. We provide a mask splitting algorithm which can separate the concave masks. Based on the evaluation of 40 neural network proposed masks of medical images, the metric properties we fine tuned.

Conclusion: We implemented a new segmentation evaluation metric [1], and we provided a pipeline which can be easily used for automatic measurement of clinical relevance of medical image segmentation.

The scientific work was reached with the sponsorship of Gedeon Richter Talentum Foundation in framework of Gedeon Richter Excellence PhD Scholarship of Gedeon Richter.

[1] Kim, Haksoo and Monroe, James I and Lo, Simon and Yao, Min and Harari, Paul M and Machtay, Mitchell and Sohn, Jason W, Quantitative evaluation of image segmentation incorporating medical consideration functions, *Medical physics* **42** (2015), 3013–3023.