Improving the Random Walker algorithm for interactive 3D medical image segmentation using AI predictions

Modify the weight function to rely on an ensemble of segmentation predictions

Author: Bram Stellinga
Supervisors: Klaus Hildebrandt, Nicolas Chaves-de-Plaza
EEMCS, Delft University of Technology

1.Introduction

The segmentation of anatomical structures in 3D medical images is often performed using Active Learning (AL), due to the limited accuracy of automated algorithms. As proposed by [1], segmenting a 3D image using AL is an interactive process where a user iteratively improves the algorithm by annotating initial labels aswell as additional labels.

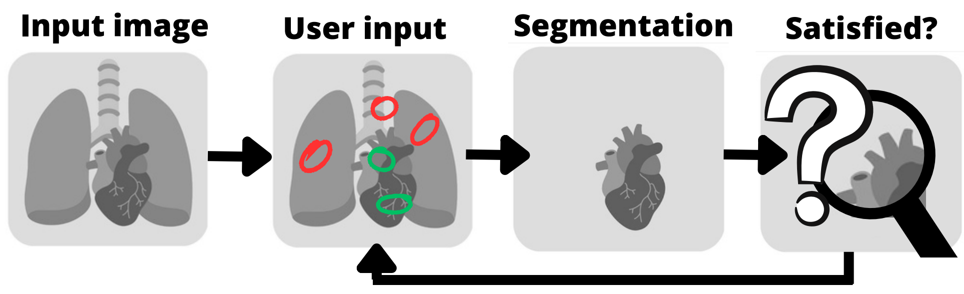


Figure 1: Simplified illustration of interactive image segmentation

2. Research Question

A Random Walker (RW) [2] is used by [1] as the segmentation algorithm. In this research, we will replicate [1] and combine the input data with an ensemble of segmentation predictions obtained by a trained Bayesian Deep Neural Network [3]. This research aims to answer the following question:

How can the performance of the Random Walker for interactive 3D image segmentation be improved by integrating an ensemble of AI-based segmentation predictions?

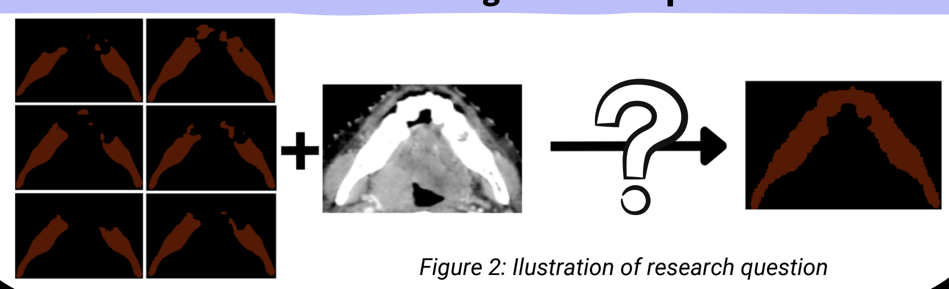


Figure 2: Ilustration of research question

5. Discussion

- All methods performed the **worst** on **optical organs**. This is likely because these are the smallest organs and thus there is limited information and more sensibility for artifacts.
- α = 0 **outperforms** all other values for α. This is likely because seedpoints are far away from most voxels, as 2 iterations only provide little user input and the original weight function **decreases faster** than prediction based, as image intensities have more variance overall.
- Unanimous Vote performs **better** than the Mean Prediction. This is likely because by minimizing the influence of non-unanimous votes it **reduces sensitivity** to prediction variances.
- Adaptive Alpha **outperforms** all methods. This makes sense as by summing the data beforehand, it gives a representation of edges/boundaries in the image depending on both the predictions and intensities.

3.Methodology

Random Walker

The Random Walker works by assigning labels to voxels based on their likelyhood to arrive at a labeled seed point. This likelyhood is defined by a weight function $w_{i,j}$ that indicates weights between neighbouring voxels i and j , based on their similarity. For the original RW this is based on **intensity.**

Integrate AI predictions

We have an ensemble P consisting of N predictions, each labeling a voxel in the image, either 1 or 0. To modify the weight function to integrate P, we compare **4 methods:**

1. Original RW:
- Based on similarity in **intensities:** $(I_i - I_j)^2$

2. Mean Prediction:
- Based on similarity in **average prediction:** $(P_{mean,i} - P_{mean,j})^2$

3. Unanimous Vote:
- Based on similarity in **unanimous prediction:** $0.5 \;\; if \;\; 0 > P_{mean,i} < 1 \vee 0 > P_{mean,j} <$ $|P_{mean,i} - P_{mean,j}| \;\; else$

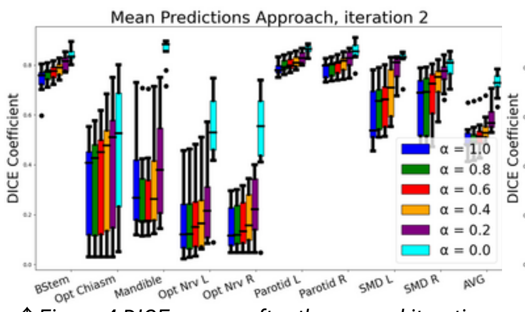
4. Adaptive Alpha:
- Based on similarity in **average prediction + intensities:** $(S_i - S_j)^2$ $S = \frac{P_{mean}}{\sigma_{P_{mean}}} + \frac{I}{\sigma_I}$

Figure 3: The 4 methods presented in this research. I_i corresponds to the intensity of voxel i and P_mean,i the mean prediction for voxel i. σ represents the standard deviation.

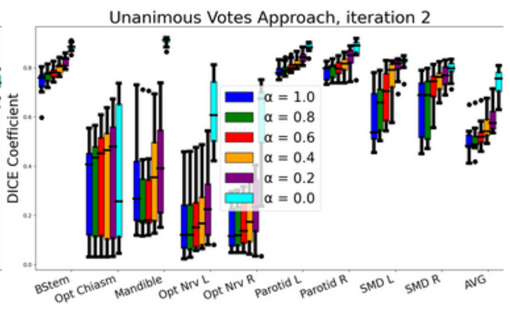
Lastly, methods **2** and **3** will be evaluated as a **weighted sum** with method **1** for some α: $w = \alpha \cdot w_{original} + (1 - \alpha) \cdot w_{new}$

4. Results

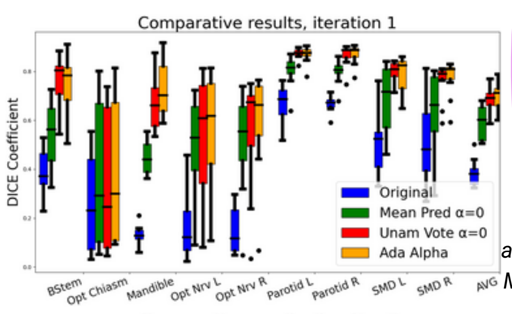
The following results are obtained by testing on a MICCAI dataset [4]. Each figure shows boxplots indicating the average DICE Coefficients of 10 patients, for 9 different organs. The last position on the x-axis shows the average of all organs.



↑ *Figure 4:DICE scores after the second iteration using Mean Prediction method for different α.*



↑ *Figure 5: DICE scores after the second iteration using Unanimous Vote method for different α.*



Both fig 4 and 5 show:
- α = 0.0 performs the best
- Correlation between α and score seems exponential
- Optics produce worst results

←*Figure 6: DICE scores after first iteration (top) and second (bottom). This figure shows results of Mean Prediction, Unanimous Vote (both using α = 0), the Original RW and Adaptive Alpha.*

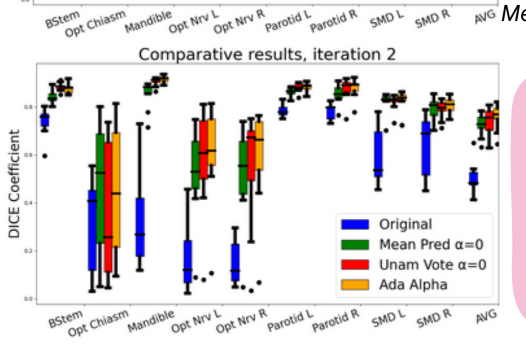


Fig 6 shows:
- Adaptive Alpha performs the best
- Unanimous Vote outperforms Mean Prediction
- All better than Original RW
- All perform worst for optics

6. Conclusion & Future Work

The **main findings** are:

- By minimizing the dependence on non-unanimous predictions it performs **better** than considering only the average predictions
- When combining normalized average prediction with the image intensities beforehand, it **outperforms** all other methods including the original RW.
- Summing the different weight functions afterwards with the original weights does **not** increase performance.

For **future work**:

- It would be beneficial to test on **less accurate** predictions.
- Additionally, it would be beneficial to define an **indication of certainties** for each voxel, both for the predictions and the image intensities. This could be based on for example the standard deviation.
- Lastly, to make the certainty values less sensitive for outliers, it is a desired future work to only take into account values in a **local neighbourhood** around each voxel.

References

[1] Andrew Top, Ghassan Hamarneh, and Rafeef Abugharbieh. Active learning for interactive 3d image segmentation. In Gabor Fichtinger, Anne Martel, and Terry Peters, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, pages 603–610, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[2] L. Grady. Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(11):1768–1783, 2006.

[3] Mody, P.P., de Plaza, N.C., Hildebrandt, K., van Egmond, R., de Ridder, H., Staring, M.: Comparing Bayesian models for organ contouring in head and neck radiotherapy. In: Colliot, O., I˘sgum, I. (eds.) Medical Imaging 2022: Image Processing. vol. 12032, p. 120320F. International Society for Optics and Photonics, SPIE (2022). https://doi.org/10.1117/12.2611083

[4] Raudaschl, P.F., Zaffino, P., Sharp, G.C., Spadea, M.F., Chen, A., Dawant, B.M., Jung, F.: Evaluation of segmentation methods on head and neck ct: Autosegmentation challenge 2015. In: AutoSeg Workshop and Challenge at MICCAI. pp. 2020–2036 (2017)