MIALab Project HS2020 - Image normalization has an important influence on the segmentation

Thomas Buchegger *University of Bern,* Carolina Duran *University of Bern,* Stefan Weber, *University of Bern*

Abstract—Here is the abstract.

January 03, 2021

I. INTRODUCTION

This is a very important citation test. [1] Postprocesses of clinical diagnosis images for treatment planning often include manual segmentation of brain regions. To segment and label the brain structures in a large dataset is complicated and time-consuming by manual operator-guided segmentation. Furthermore, it is affected by user variability and prone to limiting the standardisation. This paper will depend upon automated segmentation that segments and labels five different brain regions.

To conclude we hypothesis that normalization has an important influence in the segmentation process of brain regions. We present in this paper the acquired results applying multiple normalization methods and comparing them for every segmented brain region. Furthermore, we analyse the results and conclude our findings.

Our hypothesis: "Image normalization has an important influence on the segmentation". Aims of the introduction:

- **To demonstrate importance/impact need**
- **To demonstrate novelty**
- **To justify the hypothesis / aims / investigated technology**
- **To establish expectations/scope of the report**

Some questions which could be answered: Currently, z score normalization is implemented.

- **Are there more powerful normalization methods?**
- **Is normalization really needed?**

T. Buchegger, C. Duran and S. Weber are students from the University of Berne in Switzerland.
E-mail of T. Buchegger: thomas.buchegger@students.unibe.ch
E-mail of C. Duran: carolina.duran@students.unibe.ch
E-mail of S. Weber: stefan.weber1@students.unibe.ch

- **Is the provided data already normalized in some way?**
- **Can you also unnormalize data to show a negative effect?**

For the research:

- **Understand the problem**
- **What have others already done / is there already a solution?**
- **Can I apply a solution to another problem to my problem?**
- **What lessons can I learn from others work?**
- **What deficiencies exist in others work?**

1) **Demonstrate importance**
   a) **Define the problem**
   b) **Explain the criticality, impact of the problem**

2) **Demonstrate Novelty**
   a) **Explain what is the state of the art / current practice**
   b) **Explain what preliminary / related work has been done towards solving the problem by you and others**
   c) **Explain what deficiencies / problems still exist (specifically the one that you will try to address)**

3) **Present and Justify the Hypothesis / aim / objective**
   a) **State Hypothesis /aim/ objective**
   b) **Describe evidence supporting hypothesis**

4) **Establish expectations of the report**
   a) **Describe scope of the presented work**
   b) **Present a summary of remainder of the report**

II. MATERIALS AND METHODS

MIA pipeline in general (short overview), your experiment in detail

A. Medical Backgrund

The five brain regions segmented in this paper are the white and grey matter, hippocampus, amygdala and the thalamus. All regions are visible in figure **??**.
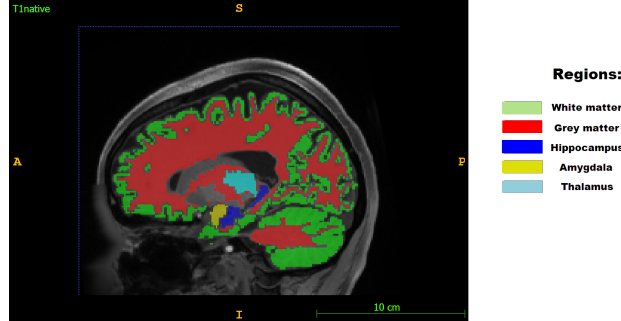

Fig. 1: All brain regions

B. Data

The used data were 30 unrelated healthy subjects from the Human Connectome Project data set. Overall, the dataset consisted of 30 MRI patient images, out of which 20 were used for training and 10 for testing the model. From each patient a 3 tesla T1- and a T2-weighted image (T1w and T2w image) was available. Each MRI file was of the size 118x118x217 pixels. All images are skull defaced for anonymization. Furthermore, the corresponding brain masks and the ground truth segmentation exist for all subjects. An atlas in MNI152 space is also available.

C. Registration

To register the T1w and T2w image data an affine transformation was applied. The transformation was found by an intersubject registration from the T1w image to the provided atlas. The corresponding transformations have already been determined and are not part of this work.

D. Preprocessing

As already stated the aim of this work is to determine the influence of the preprocessing and especially of the normalization on the performance of the segmentation. Therefore in total six different normalizations were applied to the image data. And the results were compared among all normalizations and no normalization. As an additional preprocessing step skull stripping was performed before normalization on all runs with the brain mask provided. To provide some context all used normalization techniques are briefly explained in this section. The used normalizations already established and for additional information can be found.

The following normalization techniques were used:

1) ZScore: Normalize an image by subtracting the mean intensity and dividing by the standard deviation of the input image intensity. This transforms the image data into an intensity distribution with a mean of 0 and a standard deviation of 1.

$$I_{New} = \frac{I - \mu}{\sigma} \tag{1}$$

2) MinMax: Normalize an image by subtracting the minimal intensity value and dividing by the difference of the maximal and minimal intensities. This scales the intensities in a range from 0 to 1.

$$I_{New} = \frac{I - I_{min}}{I_{max} - I_{min}} \tag{2}$$

3) Whitestripe: In the whitestripe normalization also formula (1) is used. In contrast to the Z-Score normalization μ and σ obtained from the intensity values of normal appearing white matter (NAWM). After smoothing the histogram NAWM is found by selecting the highest intensity peak of the T1-w image. μ corresponds to this peak. Around the obtained μ a subset of 10 % from all pixel values is taken and the standard deviation σ is calculated from this subset. This subset is called the whitestripe. By applying formula (1) with the found μ and σ the peak of the white matter is shifted to 0 and the intensities are scaled with σ.

4) Fuzzy-C Means: In this normalization a mask of the white matter is crated by using the Fuzzy C-means algorithm. The found white matter tissue mask is used to calculate the mean μ of all intensities which correspond to the white matter. Then all image intensities are scaled by μ and shifted to a constant target value c

$$I_{New} = \frac{c \cdot I}{\mu} \tag{3}$$

5) Gaussian Mixture Model:
6) Histogram Matching:

E. Feature Extraction & Classifier

The classifier used was a Random Forest classifier with which different parameters with all used normalization techniques were combined. Which parameters lead to which results can be found in the results section.

F. Postprocessesing

G. Evaluation

For this project, the used normalization techniques were evaluated on the T1- and T2-weighted MRI data sets. For this evaluation the Dice Similarity Coefficient (DSC) as well as the Hausdorff distance was chosen. DSC returns as a quality metric how much two regions overlap and is, therefore, a useful metric for segmentation. The dice coefficient of two sets, as explained in

figure 2, is a measure of their intersection scaled by their size. The result is in the range from 0 to 1, where 1 is a perfect segmentation.
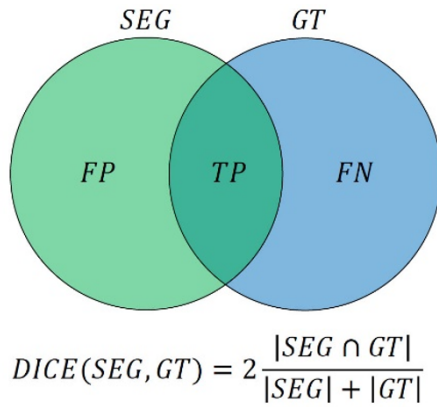

$$DICE(SEG, GT) = 2\frac{|SEG \cap GT|}{|SEG| + |GT|}$$

Fig. 2: Graphical representation of the Dice Similarity Coefficien. SEG stands for the achieved segmentation whereas GT means ground truth.

The Hausdorff distance measures how far two subsets of a metric space are from each other. So two sets are close if every point of either set is close to some point of the other set. Then the Hausdorff distance, explained in figure 4, is the longest distance from one of the two sets to the other set.

III. RESULTS

In depth analysis of experiment related results

A. Data

B. Model

C. Evaluation

IV. DISCUSSION

Aims of the Discussion part:
- **Highlight importance of your work (highlight novelty /impact etc**
- **To interpret your results in relation to your original problem**
- **To put your work into the context of existing work**
- **To present any limitations of the presented work**
- **To make future recommendations**
- **To provide a conclusion of the work**

1) **Importance of the work**
   a) **Summarise your results**
   b) **Reiterate the importance of the work (novelty , impact etc)**
2) **Interpretation of results**
   a) **Interpret your results focussing on the problem described in the introduction.**
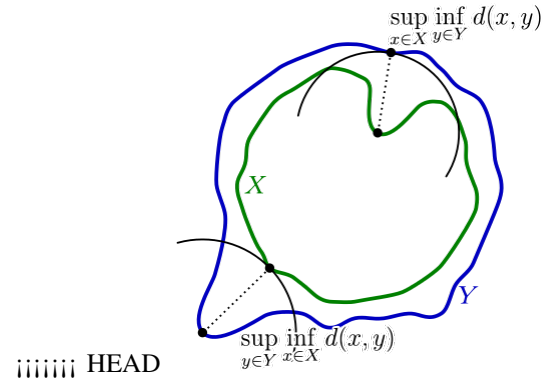

Fig. 3: Components of the calculation of the Hausdorff distance between the green line X and the blue line Y.
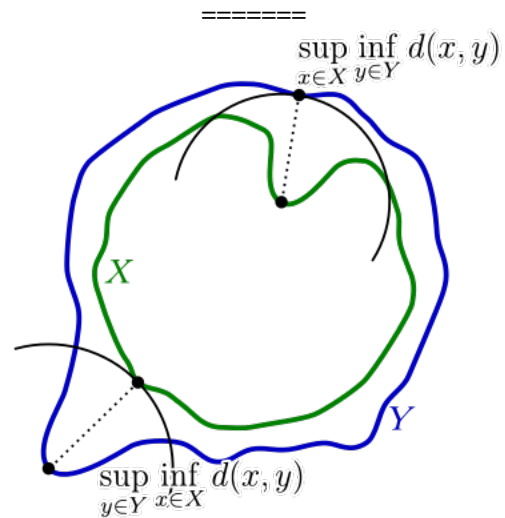
=======

Fig. 4: Components of the calculation of the Hausdorff distance between the green line X and the blue line Y.
¿¿¿¿¿¿¿ f368b2b55c36fdd80c03e7e927c60bab5ef27044

   What do the results mean for the described problem?
   b) **Explain any unusual/important findings (be careful if not your original investigative subject)**
3) **Provide context**
   a) **Describe your results in relation to others and try to explain any discrepancies**
   b) **Emphasize how your results support or refute your hypotheses current thinking in the field. Were results as expected? If not why and what does this mean?**
4) **Limitations of your work**
   a) **Describe any limitations /deficiencies of your work and what impact they have on the findings**
   b) **Suggest possible future solutions**

V. Conclusion

References

[1] Y. H. Roohani and E. G. Kiss, "Improving Accuracy of Nuclei Segmentation by Reducing Histological Image Variability," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11039 LNCS, pp. 3–10, 2018.

- **Summarise your findings and relate your findings back to your hypothesis / aim / objective and to your problem.**
- **Based on your findings, suggest next steps towards solving your problem**