A Client/Server based Online Environment for the Calculation of Medical Segmentation Scores

Maximilian Weber, Daniel Wild, Jürgen Wallner, Jan Egger

Abstract—Image segmentation plays a major role in medical imaging. Especially in radiology, the detection and development of tumors and other diseases can be supported by image segmentation applications. Tools that provide image segmentation and calculation of segmentation scores are not available at any time for every device due to the size and scope of functionalities they offer. These tools need huge periodic updates and do not properly work on old or weak systems. However, medical use-cases often require fast and accurate results. A complex and slow software can lead to additional stress and thus unnecessary errors. The aim of this contribution is the development of a cross-platform tool for medical segmentation use-cases. The goal is a device-independent and always available possibility for medical imaging including manual segmentation and metric calculation. The result is Studienfenster (studienfenster.at), a web-tool for manual segmentation and segmentation metric calculation. In this contribution, the focus lies on the segmentation metric calculation part of the tool. It provides the functionalities of calculating directed and undirected Hausdorff Distance (HD) and Dice Similarity Coefficient (DSC) scores for two uploaded volumes, filtering for specific values, searching for specific values in the calculated metrics and exporting filtered metric lists in different file formats.

I. INTRODUCTION

Medical image processing covers a wide range of applications [1]. An important part of these applications is medical image segmentation [2]-[4]. Although there are many different approaches in semi-automatic and automatic segmentation algorithms, with Deep Learning [5] achieving currently the best results, manual segmentation remains the priority due to the accuracy in most medical segmentation applications [6], [7]. Additionally, medical image segmentation algorithms are needed to evaluate and compare segmentations. These algorithms are used and proven for years in all different fields of image processing.

Although image segmentation has grown rapidly in medicine a major part of the tools and applications stayed the same for years. Especially in terms of availability, cross-platform support and usability there is a major room for improvements. This contribution aims to remedy the mentioned problems by the development of a cross-platform web-tool for manual image segmentation and calculation of segmentation scores.

There are two major use-cases aimed by this tool. The first is the need for a manual segmentation software that is always and everywhere available. Widely used desktop applications for medical image segmentation are often very large due to the broadly ranged possibilities offered for all areas of medical image processing. Often huge updates are needed, which require nightly builds, even if only the functionality of segmentation is needed by the user. Examples are 3D Slicer [8], MeVisLab [9] and MIPK [10], [11], which have installers that are already over 1 GB. In addition, in sensitive environments, like hospitals, it is mostly not allowed at all to install these kind of open-source software tools, because of security issues and the potential risk that patient records and data is leaked to the outside.

The second is the need of software for easy and stable manual segmentation training. A lot of freeware and open-source segmentation tools are not very user-friendly and buggy. In contrast, commercial medical imaging processing tools, which is more stable, is often very expensive. Finally, deploying the software to every medical student or segmentation trainee is cost intensive.

With the tool developed within this contribution, manual image segmentation is always available for everyone. Experts but also trainees can be supported by the tool. There is no need of deploying the software to each user and to update it regularly. As a user-study has shown, the manual segmentation and the calculation of the segmentation metrics can be accomplished in an efficient and user-friendly way within the developed tool. The user-study has also proven that even attendees without any medical and image processing background were able to segment a brain tumor with considerable accuracy after a short introduction.

II. METHODS

A. Medical Use Case

Segmentation analysis is very important in order to show the development of special disease in the human body. Another scope of application is the analysis and evaluation of semi-automatic or automatic segmentation algorithms.

For both use cases, a ground truth is needed. In the first use case, the disease development, we have two states that have to be compared: a "first" state of the disease, which is compared to the disease after a certain time. Due to this comparison, it is possible to recognize if the disease treatment has helped. To perform the comparison an image registration and segmentation of the two images with the disease need to be done first.

J. Egger is with the Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, 8010 Graz, Austria (phone: +43 316 873 5076; fax: +43 316 873 5050; e-mail: j.egger@icg.tugraz.at).

M. Weber and D. Wild are with the Institute for Computer Graphics and Vision, Graz University of Technology, Austria.

J. Egger and J. Wallner are with the Department of Oral and Maxillofacial Surgery, Medical University of Graz, Auenbruggerplatz 5/1, 8036 Graz, Austria (e-mail: jan.egger@medunigraz.at and j.wallner@medunigraz.at).
In the second use case, the algorithm comparison, the ground truth is the manual segmentation done by a specialist. Based on this segmentation, the different segmentation algorithms can be compared and rated.

To get information about the accuracy of the developed tool in this contribution, a modified mixture of these two use-cases is applied. The idea is to compare the results of the Hausdorff Distance (HD) [12] and Dice Similarity Coefficient (DSC) [13] of the developed tool with the results of the HD and DSC of a widely used medical imaging tool. The data used for the comparison are mandibular Computed Tomography (CT) files which can be found at [14]. The dataset contains CT scans from ten anonymized patients. From these datasets the mandible has been segmented by two specialists. These segmentations are analyzed and compared in this contribution.

B. Setup Development Environment

The decision was made to use the open-source web-application Slice:Drop (http://slicedrop.com/) [15] as an implementation starting point. Slice:Drop already provides functionalities to load and view 3D images. These functionalities are needed for the manual segmentation part of Studierfenster, which is described in a different publication [16]. In addition, to enable the development with ITK (https://itk.org/) and VTK (https://vtk.org/), the toolkits needed to be installed on the server.

C. Implementation of Use Case

Before starting the implementation, a concept was needed, how the implementation should be accomplished. The handling and usability of the tool should be very simple and the tool should be usable for the day-to-day work.

At first, the decision was made to develop a web tool to ensure multi-platform usage. Since the calculation of the segmentation scores cannot be done client-sided, the development splits into a client and a server part. A concept of how the communication between client and server looks like when performing an image segmentation analysis is shown in figure 1.

![Client-Server Communication](image)

**Figure 1.** Client-Server Communication.

D. Implementation of the Client

The design focus of the landing page was a clear layout. The functionalities should be visible at first sight. Four functionalities are included on the page. The first one is the HD and DSC calculation, which is addressed in this contribution. Additional functionalities of Studierfenster are a converter of mesh files, to convert different file-formats to the NRRD file-format, which is accepted by the tool, the already mentioned manual segmentation and a Virtual Reality (VR) viewer [17]. The NRRD file-format was chosen, because it has the benefit, that medical images get anonymized. The segmentation score calculator and the converter are opened as a pop-up window within the webpage. The medical image visualization and segmentation view and also the VR view are opened as new windows. Figure 2 presents the current landing page of the Studierfenster web-application.

![Studierfenster Landing Page](image)

**Figure 2.** The current landing page of the Studierfenster tool.

Due to the fact that the tool should be considered as a standalone tool, it was necessary to clearly distinguish the design from Slice:Drop and other tools in this scientific area.

To consolidate the unique characteristic a logo design and additionally an outstanding name was needed. The decision was made to call the tool 'Studierfenster' following the Studierstube software framework, which has been developed at our institute [18], [19]. The logo should symbolize a brain and gears to show the scope of application, which covers different medical image processing use-cases.

One important feature of the tool is the segmentation analysis functionality. One should have the possibility to calculate HD and DSC values out of two medical image volumes directly on the webpage.

The next parts of the implementation were done client-sided. In this contribution, the following functionalities within the segmentation analysis were implemented: an uploading interface with the possibility to upload two files at a time, a calculation output, the possibility of a tabular view of different results, search functionality through the different calculations in the tables and different export possibilities for the tables.

Figure 3 presents the segmentation analysis window without any uploaded files and calculations. This window can simply be closed on the x-sign on the right top and will show the landing page again.

![Segmentation Analysis Window](image)

**Figure 3.** The segmentation analysis window.
The whole table can be deleted when pressing the button 'Clear Table'. This should provide a more fluid workflow within the tool.

E. Implementation of the Server

As Server environment the Python framework Flask (http://flask.pocoo.org/) is used. In order to process the AJAX requests sent from the client to the server, the server needs information where the requests will arrive and how the requests should be handled. Each request needs to be routed to a specific URL. These URL's will then be observed by the server. With the AJAX request an 'id' and also file information can be transmitted. The 'id' can be used as transfer value.

One request needed is the file upload. If the file upload is not successful, the server returns an error to the client. For the calculations, information about the file names and the information if image spacing should be used is needed. These information are also transmitted to the server via the requests. On the server, three different calculations are performed. These calculations are done by python scripts. For the simultaneous execution of the calculations, subprocesses are started on the server. The output of the subprocesses are piped to the main process to receive the output of the Python scripts. On success, the Python output is returned to the client. If an error occurs during the script execution, the error code is returned to the client.

After the calculation, the files are deleted from the server. The deletion can additionally be accomplished with a button click via the user, this is also done with an AJAX request, where the file information is sent to the server.

F. Implementation of the Segmentation Analysis Calculation

To get numerical values of the similarity between two volumes, the HD and DSC are calculated within the tool. The values are calculated on the server. Due to accuracy reasons also the directed HD in both directions are calculated. The calculations are triggered by AJAX requests from the client. All four calculations are implemented in Python.

For the HD calculation, ITK is used. ITK needed to be imported into the Python project in order to use it. The directed and undirected HD calculations are implemented using the 'DirectedHausdorffDistanceImageFilter' provided by ITK. For the DSC calculation, the 'LabelOverlapMeasuresImageFilter' from SimpleITK is used.

The results of all calculations are returned to the client using the piped standard output and the success message of the AJAX request.

G. User Study

To evaluate the developed tool, a user study has been done. The first part of the user study was a practical task. It had to be done within the tool without any help after a short introduction to the general functionalities. The task was split into two subtasks. First, the participants had to manually segment a glioblastoma (GBM) tumor, then they had to calculate the HD and DSC. The dataset was provided by the university hospital of Giessen, Germany, and contains a Magnetic Resonance Imaging (MRI) scan of a GBM. The results of the segmentations were compared with the ground truth.
segmentation of an expert, who had also used the developed tool for the task.

The second part of the user study was a questionnaire including questions about the user interface of the tool, the usability of the tool and the general experiences performing the tasks with the tool.

Participants of the user study were from different expertise areas. Half of them had a medical background (from medicine students to senior surgeons), the other half had no medical background (mostly law and computer science students). The questions used for the user study questionnaire were derived from the questionnaire questions from [20]. The questions are ISONORM 9241/10 compliant.

III. RESULTS

In order to compare the calculation results between the tool developed within this contribution and MeVisLab, datasets are needed. For the segmentation analysis calculations, datasets from [21] were used. These datasets contain segmentations of ten patients’ mandibles from CT scans. The segmentations were done manually by two facial surgeons (A and B).

In table 1, the results of the segmentation analysis calculations are presented. The DSC and HD were calculated in the developed tool and as a comparison in MeVisLab. For the HD the results for the calculation with image spacing (IS true) as well as the results for the calculation without image spacing (IS false) are shown.

| File 1 | File 2 | Studierfeuster DSC IS true HD IS false DSC IS true HD IS false |
|--------|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Pat 1 Seg A | Pat 1 Seg B | 0.9468 | 1.0286 | 3.1621 | 0.9433 | 1.0286 | 3.1623 |
| Pat 2 Seg A | Pat 2 Seg B | 0.9176 | 2.1892 | 3.1623 | 0.9176 | 2.1892 | 3.1623 |
| Pat 3 Seg A | Pat 3 Seg B | 0.9265 | 3.0189 | 5.1062 | 0.9265 | 3.0189 | 5.1062 |
| Pat 4 Seg A | Pat 4 Seg B | 0.9468 | 2.7013 | 3.1623 | 0.9468 | 2.7013 | 3.1623 |
| Pat 5 Seg A | Pat 5 Seg B | 0.9689 | 2.0924 | 6.3246 | 0.9688 | 2.0924 | 6.3246 |
| Pat 6 Seg A | Pat 6 Seg B | 0.9418 | 2.2388 | 3.3166 | 0.9418 | 2.2388 | 3.3166 |
| Pat 7 Seg A | Pat 7 Seg B | 0.9111 | 2.6995 | 4.1231 | 0.9111 | 2.6995 | 4.1231 |
| Pat 8 Seg A | Pat 8 Seg B | 0.9461 | 2.4600 | 4.1231 | 0.9461 | 2.4600 | 4.1231 |
| Pat 9 Seg A | Pat 9 Seg B | 0.9041 | 1.6668 | 2.2061 | 0.9041 | 1.6668 | 2.2061 |
| Pat 10 Seg A | Pat 10 Seg B | 0.9555 | 1.5 | 3.3166 | 0.9555 | 1.5 | 3.3166 |

Table 1. Results of the segmentation analysis calculations.

Regarding the calculation results, the developed tool delivers the exact same metric results as widely used desktop applications. This is a basic requirement and necessary for a practical application. These results were expected.

A major advantage compared to other applications is the immediate possibility to filter and search through the metric calculation results and also the functionality of exporting the filtered metrics into different file formats.

This advantage may compensate for the fact, that the calculation times with the tool are sometimes longer compared to offline desktop applications. Due to the fact, that a typical segmentation and metric calculation workflow includes calculations of more than one file, the easy summary of the results can also be beneficial for practical applications. Another difference to offline applications is the need of uploading the medical images (segmentation masks) in order to calculate the HD and DSC metrics. For large volume files, this may take some time, but with an average internet speed, the uploading time is negligible. Because of the NRRD file format, the metadata of the medical data is automatically anonymized, which is necessary for critical personal information. In addition to the anonymization, the uploaded files are deleted from the server immediately after the calculation has been performed.

The first oral feedback received from the participants was consistently positive. The goal was to provide a relaxed yet professional environment for the user-study so that the participants are not stressed, but still take the study seriously. The training period was accepted very well. The average time spent for explanation and the pre-training was five minutes.

After the training, the participants started the first task, the manual segmentation. The segmentation worked very well for all participants without interruptions or inconsistencies. The time spend for the segmentation task ranged from three minutes to 15 minutes. The average time spent by the medical group was nearly nine minutes while the average time spent by the non-medical group was around six minutes. The difference in spend time can be explained by the fact that medical participants did slightly more segmentation adjustments than non-medical participants.

After the segmentation task, the participants had to calculate the DSC and HD for their segmentations with the medical score calculator presented within this contribution. The calculation also went well for all participants. There are clear differences between the medical and non-medical group in the segmentation metrics. The DSC range for medical participants is between 83% and 87% with an average value of 85% and for non-medical participants between 82% and 84% with an average value of 84%. A high DSC value means, that the volume is very similar to the ground truth. The HD range for medical participants reaches from 2.8 voxels to 4.9 voxels with an average distance of 3.5 voxels. The range for non-medical participants reaches from 3.2 voxels to 4.8 voxels with an average distance of 3.8 voxels. For HD lower numbers are better than higher ones. The results for HD also show some differences between the two groups, although the difference is lower.

Even if the medical group had slightly better results for DSC and HD, the results of the non-medical group were also very good, considering the fact that they saw a brain tumor in a MRI scan for their first time and never segmented something like this before.

The last part of the user study was the questionnaire. The range of the questionnaire was eight questions with a point scale from zero (worst) to seven (best). Even if the results are overall very satisfying, some findings can be observed from the answers. Three questions had the minimum average rating of 5.7 points. First of them was a question about the training time, the second was a question about the availability of all necessary functions to achieve the goal and third was the general opinion about the user interface. With the first and third of these questions, the calculation window was targeted. The second was about the manual segmentation. All three mentioned topics offer some room for improvements. For example, the graphical interface of the metric calculation window could have a more intuitive design. In addition, a few more functionalities of the manual segmentation could be added.
The questions with the next higher average result of 6.0 and 6.2 points, address the satisfaction of the segmentation result. Improving the first three topics would lead to better segmentation results overall and therefore would also improve this topic.

The last three questions have nearly the same amount of average points. The questions are about the complexity of the software, the satisfaction of the needed time for the segmentation process and the overall impression, which got a very good rating of 6.4 and 6.5 points.

IV. CONCLUSION

Taking all the results into account, the targeted goal of this contribution was reached. A platform independent tool for manual segmentation and calculation of segmentation scores has been developed and evaluated. For basic tasks and training the tool is already ready for usage and can be accessed worldwide via a standard web browser from a desktop PC, smartphone or tablet. In addition, a user study with participants with different backgrounds achieved very good results even after a short training time and with little previous knowledge.

There are several areas of future work, for example, providing more ITK and VTK functionalities within our tool to the end user [22]. This could be advanced segmentation and registration algorithms that are already implemented within these libraries, like ITK-SNAP [23]. Furthermore, incorporating other research groups and collaborators to add more functionalities to Studierfenster, thus providing their research findings to a broader audience. Finally, enabling the creating and training of Deep Learning networks via our web environment [24]-[26] and an interface to Augmented Reality (AR) head mounted displays [27], like the HoloLens [28].

ACKNOWLEDGMENT

The work received funding from the Austrian Science Fund (FWF) KLI 678-B31: “enFaced: Virtual and Augmented Reality Training and Navigation Module for 3D-Printed Facial Defect Reconstructions” (Principal Investigators: Drs. Jan Egger & Jürgen Wallner), the TU Graz Lead Project (Mechanics, Modeling and Simulation of Aortic Dissection) and was supported by CAMed (COMET-K-Project 871132), which is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian Federal Ministry for Digital and Economic Affairs (BMDW) and the Styrian Business Promotion Agency (SFG).

REFERENCES

[1] G. Dougherty, “Medical Image Processing: Techniques and Applications (Biological and Medical Physics, Biomedical Engineering),” 2011