Crowdsourcing ground truth data for analysing brainstem tumors in children

Benjamin Timmermans Zoltán Szlávik Robert-Jan Sips

Vrije Universiteit, Amsterdam

Abstract

Brainstem tumors are a rare form of childhood cancer for which there is currently no cure. The Semmy Foundation and IBM Netherlands are developing a cognitive system for quicker analysis of MRI-scans and better detection of anomalies in the brainstem. In order to gather training data for this system, the Semmy Annotation Tool was developed that allows untrained participants to draw the exact shape of the brainstem and tumor onto an MRI-scan. The tool was tested in an experiment at the Lowlands 2016 festival, in which 823 participants annotated 5.152 images.

1 Purpose

Diffuse intrinsic pontine glioma is a rare type of tumor located in the center of the brainstem. Yearly around 18 children die in The Netherlands through this disease, and there is currently no known cure. The Semmy Foundation[1] was founded with the goal of increasing the survival chance of children with this type of cancer. Building on the work of Kaspers et al. [2], the Semmy Foundation has started a new initiative together with the Center for Advanced Studies at IBM Netherlands to develop a cognitive system that should allow doctors and researchers to quicker analyse MRI-scans and better detect anomalies in the brainstem. In order to train this system large amount of annotated data is required. The Semmy Annotation Tool was developed in order to gather this data through crowdsourcing.

The Semmy Annotation Tool[2] was developed and tested at the science fair of the festival Lowlands 2016. The participation in this fair allowed: 1) to increase the awareness for the Semmy Foundation and this type of cancer, 2) to gather large amounts of annotations on the shape of the brainstem and cancer cells, and 3) to measure the quality of the annotations that a participant made in relation to the well being of that participant. The hypothesis is that people under influence can still make valuable contributions, but that these are of lower quality than sober people. This study may give insight into the reliability of online crowd workers and the effects of various aspects on the annotation quality. This is a problem in crowdsourcing that to our knowledge has not yet been investigated.

2 Methods

The dataset used for the experiment contained 1.711 anonymized images of MRI-scans on children with brainstem cancer, which were made available by the Semmy Foundation. The Semmy Annotation Tool was developed so multiple untrained participants could simultaneously draw the edges of the brainstem precisely on these medical images in a controlled environment. The tool uses the Laravel PHP framework, and Semantic UI for the interface. When an image is shown for annotation it is scaled up to the maximum size within the canvas, after which drawing of the shape on an image is made possible through custom javascript code (See Figure 1). This differs from approaches like [1], where the placing of independent polygons is used and the assumption is made that the annotator can always identify the shape that is to be found.

[1] http://www.stichtingsemmy.nl/

[2] http://github.com/CrowdTruth/Semmy/

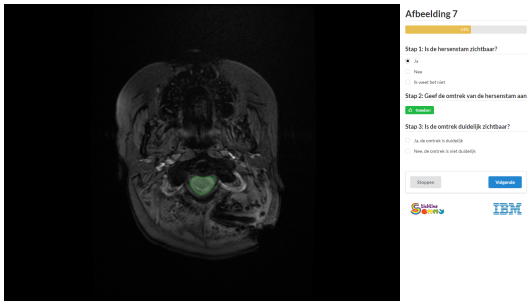

Figure 1: Semmy Annotation Tool used to draw the shape of the brainstem and tumor.

Figure 2: Dashboard used to motivate the crowd to participate.

In the crowdsourcing task[3] the participant was first asked for age, gender, education level, and whether alcohol or drugs had been used. This should allow the precision of the annotations to be matched to the registered well being of the participant. In order to guarantee correctness of this input, moderators were assigned to assist and give the initial instructions for the annotation task. After this the participant was given two minutes to annotate images. The images were ordered randomly, but with priority for images with the fewest annotations yet. For each image the participant was asked whether the brainstem or tumor could be seen. If the answer was yes, the participant was asked to draw the shape onto the image, and to indicate whether the edges were clearly visible in the image. Based on the CrowdTruth methodology [3] for capturing human interpretation, the goal was to gather 10 to 15 annotations for images where the tumor was indicated to be seen.

In order to motivate the audience to participate, a dashboard[3] was created (Figure 2) that shows live aggregated statistics such as the amount of participants and the amount of annotations gathered. Also, a 3D visualization of the brainstem was added to the dashboard in order to increase the understanding of what the brainstem looks like. In the experimental setup at Lowlands the participants could annotate simultaneously on two laptops and the dashboard was shown on a large tv screen.

3 Results

In total 823 people participated with an average age of 28.3 years. Of the participants 58% were female, and 69% had finished at least a bachelors degree. In total they annotated 5.152 images, of which 2.440 were said to contain a brainstem or tumor. This indicates almost half of the images were either not clear or did not contain parts of the brainstem. Within the images that were said to contain the brainstem 558.887 polygons were drawn, resulting in an average resolution of 230 polygons per image. The developed tool proved useful for gathering the data, but more annotations are needed for further analysis of the annotation quality and classification of brainstem cancer.

References

[1] Alberto Chávez-Aragón, Won-Sook Lee, and Aseem Vyas. A crowdsourcing web platform-hip joint segmentation by non-expert contributors. In *Medical Measurements and Applications Proceedings (MeMeA), 2013 IEEE International Symposium on*, pages 350–354. IEEE, 2013.

[2] GJ Kaspers. A twenty-year review of diagnosing and treating children with diffuse intrinsic pontine glioma in the netherlands. *Diffuse Intrinsic Pontine Glioma Clinical Aspects and Imaging*, 1:41, 2014.

[3] O. Inel et al. CrowdTruth: machine-human computation framework for harnessing disagreement in gathering annotated data. In *The Semantic Web–ISWC*. 2014.

[3] Screencast available at: `https://youtu.be/8VhZ7SGn0IE`