Semi-automatic Brain Tumor Segmentation by Drawing Long Axes on Multi-plane Reformat

David Gering[✉], Kay Sun, Aaron Avery, Roger Chylla,
Ajeet Vivekanandan, Lisa Kohli, Haley Knapp, Brad Paschke,
Brett Young-Moxon, Nik King, and Thomas Mackie

HealthMyne, Madison, WI 53717, USA
david.gering@healthmyne.com

Abstract. A semi-automatic image segmentation method, called SAMBAS, based on workflow familiar to clinical radiologists is described. The user initializes 3D segmentation by drawing a long axis on a multi-plane reformat (MPR). As the user draws, a 2D segmentation updates in real-time for interactive feedback. When necessary, additional long axes, short axes, or other editing operations may be drawn on one or more MPR planes. The method learns probability distributions from the drawing to perform the MPR segmentation, and in turn, it learns from the MPR segmentation to perform the 3D segmentation. As a preliminary experiment, a batch simulation was performed where long and short axes were automatically drawn on each of 285 multi-spectral MR brain scans of glioma patients in the 2018 BraTS Challenge training data. Average Dice coefficient for tumor core was 0.86, and the Hausdorff-95% distance was 4.4 mm. As another experiment, a convolution neural network was trained on the same data, and applied to the BraTS validation and test data. Its outputs, computed offline, were integrated into the interactive method. Ten volunteers used the interface on the BraTS validation and test data. On the 66 scans of the validation data, average Dice coefficient for core tumor improved from 0.76 with deep learning alone, to 0.82 as an interactive system.

Keywords: Brain tumor · Image segmentation ·
Semi-automatic · Machine learning

1 Introduction

Evidence from cancer researchers suggests that extraction of quantitative variables from medical images can contribute more information for decision support in management of cancer patients. Specifically, quantitative metrics can improve both (1) diagnostic and prognostic accuracy; as well as (2) longitudinal monitoring of patient response [1]. Criteria for monitoring radiographic brain tumor progression include the Macdonald criteria [2], Response Evaluation Criteria in Solid Tumors (RECIST) [3, 4], WHO criteria [5], and RANO criteria [6].

Currently, radiological studies are generally limited to detection and staging along with qualitative descriptions. Quantitative descriptors are not yet in the standard of care

primarily due to a lack of infrastructure and tools to derive, test, and deploy these quantitative metrics at the point-of-care for all patients. Currently available tools to do this are limited to research or clinical trials, and have not been widely deployed as they lack the speed, precision and consistency required for wider clinical use [7]. The amount of time required to delineate lesion boundaries correctly could be intrusive to the radiologist's workflow. Delineation can be performed by manually drawing the tumor boundary on each image slice, by semi-automatically guiding an algorithm, or by fully automated methods. In either the semi-automated or fully automated methods, editing is necessary. Although manual delineation offers complete control to the user, humans exhibit great variability and the process is very time-consuming. Even if an automatic or semi-automatic method were to suffer a shortcoming in accuracy, as long as there is consistency in defining the boundary, then the volume change or change in a quantitative feature may be tracked more precisely.

For MR brain tumors, recent research with fully automated segmentation, especially based on deep neural networks, has been promising [8]. SAMBAS (Semi-Automated Map-BAsed Segmentation) differs from CAD (Computer Aided Detection) because it relies on a radiologist to make an indication. The motivation is adoption by clinical radiologists who desire full-control over the segmentation, real-time feedback, an algorithm that is ready to run immediately without the need to first be trained on a large database from their site, and an algorithm whose rationale behind decisions is explainable. We expect that real-time guidance of a semi-automated approach may often have faster workflow than editing of a fully automated method.

The vital part of any measurement tool is an interface that is both familiar and effortless. Drawing the longest axis across a lesion is a natural choice for initiating contours because radiologists are already accustomed to drawing the long axis. Oncologists participating in clinical trials follow published international criteria for objectively gauging the extent and progression of disease. The Macdonald, RECIST, and WHO criteria each incorporate long axis measurements. However, inherent challenges with axis-based criteria have been reported for aggressive brain tumors [9], thus motivating the discovery of volumetric-based criteria with similar familiarity as axis-based criteria.

Besides familiarity, there are several more goals of volumetric contouring. One goal is to achieve inter-observer consistency, while also catering to individual preferences for accuracy and style. Consistency results from initialization strategies that are reproducible, such as generating 3D volumetric contours from a straight stroke rather than free-form drawing. Tailoring to individual preferences is accomplished by editing tools prepared for whenever the initial contours may be unsatisfactory. Another goal is to provide a contingency plan in case the radiologist is both unsatisfied with the contours, and unwilling to invest the requisite time to edit them. Radiologists should be given the choice of confirming either the contours (thereby enabling volumetric measures), or just the long axis, which has already been drawn, and is held in reserve as an instant alternative. Yet another goal is automatic, large-scale, quantitative validation. Given hundreds of datasets that have been manually contoured, batch processing can be implemented by calculating the long axis from each expert's contours, and employing the long axis as the simulated user input. Yet another goal is to alleviate the need to select tools from a confusing suite of options. Ideally, there is exactly one tool

in a reading room, generally applicable to all organs, yet simultaneously specialized with organ-specific features. The organ is automatically identified upon tool initialization.

SAMBAS aims to satisfy all the aforementioned goals, namely familiarity, consistency, individualism, contingency, automatic validation, and general applicability yet specialization. While advancements in processing speed have propelled deep learning (DL) in various fields, medical image analysis is missing the mass quantities of new labeled data needed for training artificial intelligence networks [10]. The multi-modal Brain Tumor Segmentation challenge (BraTS) represents a pioneering step in this direction [11, 12]. One of the goals of the software was to generate such contours on new scans, at the point of read, which in turn, can serve as the labeled image data for DL in subsequent clinical application.

2 Methods

The proposed system consists of an interactive algorithm and two compute-intensive components, which are whole-brain tissue segmentation and deep learning. Each has a run-time of roughly one minute on a typical PC, so they are run offline prior to a user's interaction with the system. While the interactive algorithm is employed by a user to segment only the core tumor, the offline components support partitioning of the tumor into its constituent parts: edema, necrosis, and actively enhancing regions.

The integration of all elements into one system is presented first, followed by the offline elements and the interactive algorithm, along with associated experiments.

2.1 System that Integrates Offline Components with User Interaction

Figure 1 presents a system flowchart. User interaction occurs in real-time because only a portion of the image is being segmented since whole-brain analysis occurred earlier.
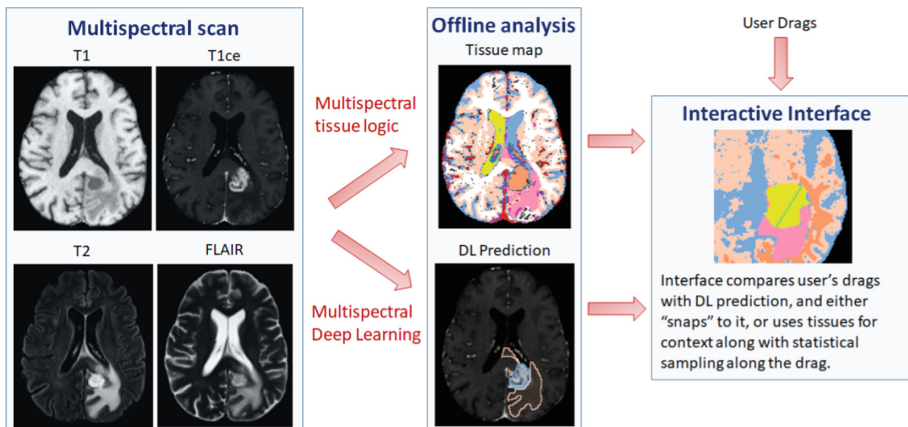

Fig. 1. Two components run offline prior to the user interacting with the system.

As the user draws a long axis on an image, only the core tumor is segmented in 2D with real-time feedback. Figure 2 displays a screenshot of the red 2D contour responding interactively to the drawn blue axis. When the user clicks a button to indicate all drawing is complete, then a 3D segmentation process runs for roughly 1–3 s, relying on the output of deep learning and tissue segmentation to find the edema associated with that particular core tumor. When a scan contains multiple distinct tumors, the user must draw a separate long axis for each tumor.
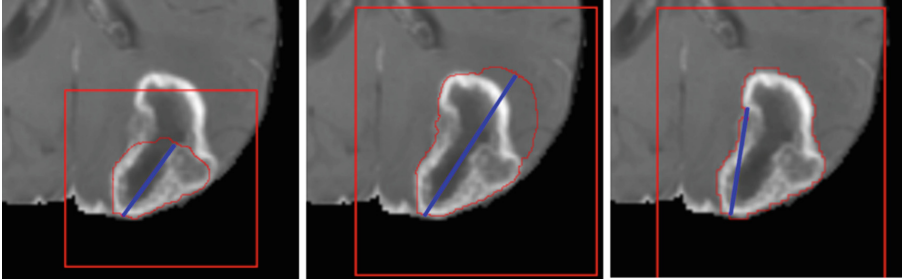

Fig. 2. These are three screenshots taken during a user's real-time interaction while drawing the long axis (blue). LEFT: The user has started drawing a long axis, but has only partially traversed the tumor at the time of the screen capture. MIDDLE: The user has over-drawn the lesion to show how the red contour always presents a reasonable result given strong image contrast in some areas, and little to none in others. RIGHT: The user has placed both endpoints of the long axis on the boundary of the output of deep learning. Consequently, the red contour "snaps to" deep learning's output contour even though the true longest axis in the plane was not indicated. (Color figure online)

As the user draws a long or short axis, whenever all endpoints of the axes are proximal to the boundary of core tumor, as found by deep learning, then the segmentation "snaps to" the output of deep learning. This process is depicted in Fig. 2. The snapping is evident to the user because a snapped contour is drawn more coarsely pixilated due to the fact that the interactive segmentation occurs on super-sampled images, whereas deep learning occurred on original images. When snapping is undesired, the user can simply hold down the CTRL key to disable it while drawing.

The 66 validation scans provided by the BraTS competition contained 89 tumors, of which 35 were "snapped to". Therefore, snapping played a role on 39% of tumors.

2.2   Whole-Brain Tissue Segmentation

The tissue segmentation classifies every brain voxel as belonging to one of several tissue types, including cerebrospinal fluid (CSF), gray matter, white matter, vessels, ventricles, and disease. Gray and white matter are found by performing Bayesian classification of the T1-weighted, contrast-enhanced image using the Expectation
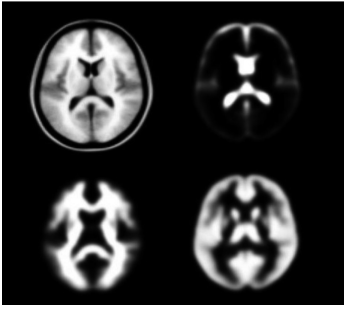
Fig. 3. The SPM atlas features an average of 305 scans (upper right) and probability maps for CSF (upper right), white matter (lower left), and gray matter (lower right).

Maximization (EM) algorithm [21]. One element of Bayesian classification is the probability that a voxel belongs to a certain tissue class prior to observing its brightness. When this prior probability varies across the image, it is referred to as a spatially-varying prior (SVP). The SVP is estimated through affine registration of the SPM atlas, as shown in Fig. 3.

Rules of logic are applied to the set of all four MR spectra to derive the other tissues. For example, enhancing tumor is described by areas that show hyper-intensity under contrast-enhancement when compared to the non-enhanced image, but also when compared to healthy white matter.

The resultant tissue segmentation will be used by the integrated system for anatomic context. For example, it will know to exclude vessels and ventricles from tumors.

2.3 Deep Learning Segmentation

Following the recent increasing successes of deep learning approaches in automated organ and tumor segmentations [13–19], a convolution neural network (CNN) was used. The CNN is based on the 3D U-Net architecture by Isensee et al. [15] (Fig. 4), which had one of the top scores in the 2017 BraTS Challenge [20]. Briefly, the input
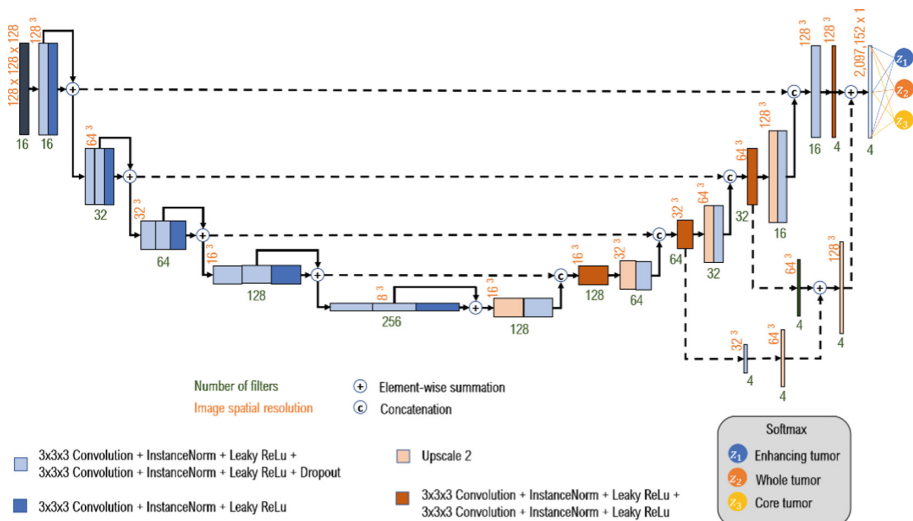

Fig. 4. 3D U-Net architecture based on Isensee et al. [3].

image data is set to $128 \times 128 \times 128$ voxels, constrained by the limited memory in the GPU. Processing from left to right, the 3D image volume is sequentially reduced in spatial resolution with multiple $3 \times 3 \times 3$ convolution layers while increasing the number of filters or feature maps as the levels move deeper. Once the lowest level is reached, the extracted feature maps are then upsampled to sequentially restore the spatial resolution at each level, concatenating with feature maps preserved during the downsampling to help restore lost information. The Softmax function classifies the 3 tumor classes. Dropouts with probability 0.3 are included to minimize overfitting.

A set of MR training data consisting of brain scans of 210 subjects with high grade glioma (HGG) and another 75 with low grade glioma (LGG) was provided by the BraTS competition. Each subject has a T1 weighted, a post-contrast T1-weighted, a T2-weighted, and a FLAIR MR image. In addition, a segmented tumor mask that contains demarcations for whole tumor, core tumor and enhancing tumor, manually demarcated by expert physicians, was also provided as the ground truth for evaluation. The images were preprocessed prior to supervised training by cropping to remove the extraneous background and preserve only the brain, resizing to $128 \times 128 \times 128$, and normalizing the MR intensities of each modality by subtracting the mean and dividing by the standard deviation.

During the supervised training, the 285 subject data were randomly split into training and validation dataset (80%–20%), each consisted of image volumes from the four different modalities and segmented truths. The training dataset was fed into the 3D U-Net model for optimization and the segmented truths were used for evaluations during the backpropagation. The model was tested at each step with the validation dataset, on which the model had not been trained. Table 1 lists the parameters used in the training of the CNN model. Batch size of one was used, despite the lower performance, in order to load the entire 285 subject image volumes into the limited memory available in the GPU.

Table 1. List of parameters used in training the CNN model.

Parameter	Value
Optimizer	Adam
Batch size	1
Initial learning rate	5×10^{-4}
Number of epochs at which learning rate is reduced	10
Learning rate reduction factor	0.5
Number of epochs at which training stopped	50

To further account for the class imbalance where there is much more background pixel data than tumor, other than cropping, a multiclass Jaccard loss function was used [14]. The four classes include 0 for background, 1 for tumor core, 2 for edema and 4 for enhancing tumor.

$$loss = -\frac{1}{K}\sum_{k \in K}\frac{\sum_i u_i^k v_i^k}{\sum_i u_i^k + \sum_i v_i^k - \sum_i u_i^k v_i^k} \tag{1}$$

The loss function is expressed in Eq. 1, where u is the prediction of the CNN and v is from the ground truth segmentation value, i is the pixel number, and k is each class in all $K = 4$ classes. The Jaccard coefficient is a measure of similarity between the segmented prediction and truth image volumes, where higher value indicates greater overlap. The multiclass version is the intersection over union of the two volumes averaged over the four classes. A negative term was added to the loss function to ensure the minimum loss function was optimized. CNN development used the open-source machine learning library, TensorFlow, and neural networking API, Keras.

2.4   Interactive MPR Segmentation

The interactive algorithm is implemented as a probabilistic framework with efficient user control. Like a digital simulation of a traditional light box on which radiologists used to view film, the 3D volume is visualized by displaying 2D planes. A Multi-Plane Reformat (MPR) refers to reformatting more than one plane, and we display a trio of planes side-by-side, such that there are axial, coronal, and sagittal orientations.

The user initializes the segmentation process by drawing a long axis on one plane of the MPR. As the user draws the long axis, a 2D segmentation updates in real-time for interactive feedback. The feedback has proven to be very helpful for the user to know precisely where to place the endpoint of the axis. Upon release of the mouse, 2D segmentation occurs immediately on the other MPR planes.

When the 2D contour is unsatisfactory, an optional short axis may be drawn perpendicular to the long axis. Other editing operations are available, such as a "ball tool" for drawing with a digital brush. A correct 2D segmentation is important since probability distributions are learned from the 2D segmentation to be employed in segmenting the other MPR planes.

When the contours on other MPR planes are unsatisfactory, then the user can draw there with the same editing tools, along with the option for drawing a long axis and short axis. This is especially useful for lesions which are irregularly shaped or oriented obliquely. Once satisfied, the user clicks a button to initiate 3D segmentation.

2.5   3D Segmentation

Multivariate Bayesian classification [21] labels image voxels as belonging to one of two classes, Background or Foreground. Classification combines the *likelihood* of class membership based on voxel brightness, with the probability of membership *prior* to observing brightness. The *likelihoods* are conditional probability distributions that do not vary across the image, while the *prior* probabilities are spatially varying, and a function of distance from region boundaries.

The user directly drives the segmentation process by manipulating four types of regions, where some regions govern the likelihoods, while some regions govern the prior probabilities. Various regions are described in Table 2, and illustrated in Fig. 5.

Table 2. Regions which drive probabilities.

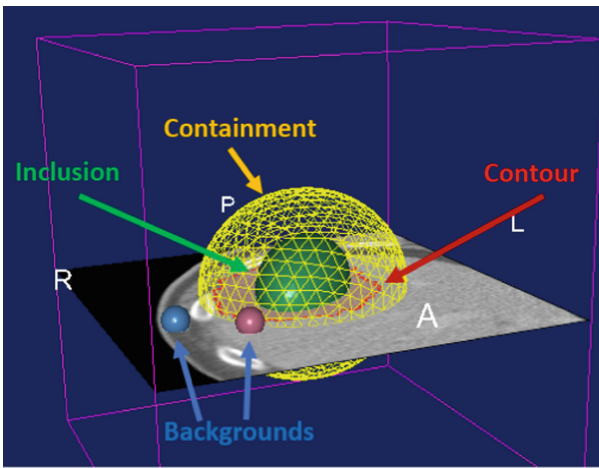| Region	Color in Fig. 1	Description
Inclusion	Green	All voxels within belong to the Foreground class, and statistically sample it
Containment	Yellow	All voxels outside belong to the Background class
Background	Blue & pink	Statistically typify Background class
Avoidance	Not shown	Spatially prohibit Foreground without affecting statistics

Fig. 5. Some regions are initially configured as ellipsoids, and then become warped. The image shown is a CT since the interactive algorithm was designed to be general purpose.

The sizes and poses of regions are automatically derived from the long axis. While the long axis describes lesion extent along one dimension, an initialization stage estimates lesion extent along other dimensions by analyzing orthogonal scout planes given statistical sampling along the long and short axes. Probability distributions are modeled parametrically as Gaussian Mixture Models (GMM) [22] while placing the Inclusion and Containment regions, and as non-parametric distributions thereafter.

Background regions are automatically placed by searching the vicinity outside the Containment region, and within the body outline, while maximizing the Mahalanobis distance [21] from the Inclusion region. Once Background and Inclusion regions are initialized, the voxels within are used to perform Parzen windowing [21] to estimate the likelihoods for Bayesian classification.

Noise and artifacts in CT vary by dose and choice of reconstruction type and kernel, and in MR by field strength, RF coil configuration, and protocol parameters, so Bayesian classification is augmented with a Markov Random Field [23] with 3 iterations of mean-field approximation.
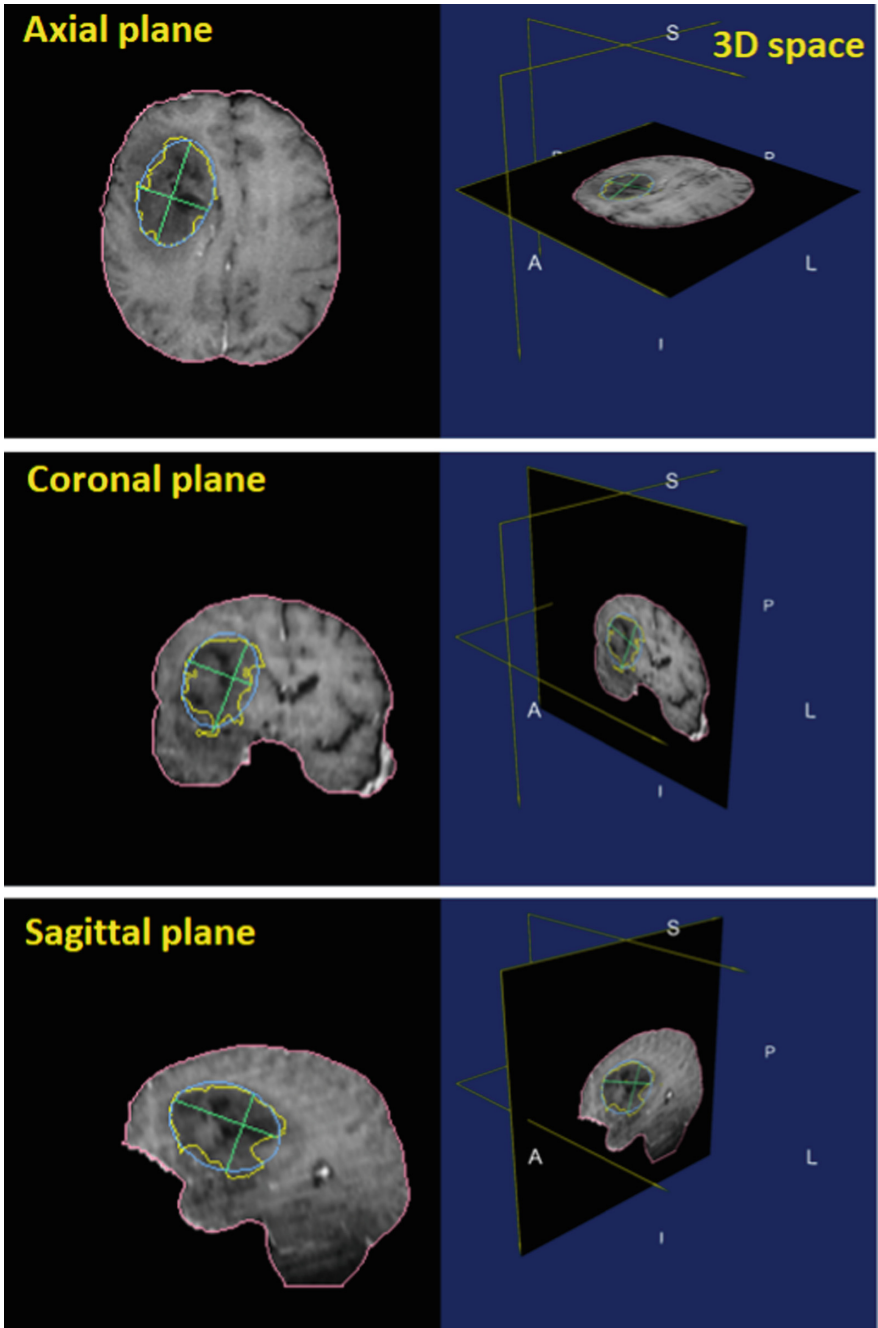
Fig. 6. Axial, coronal, and sagittal planes of MPR are shown from top to bottom. The blue ellipse was fit to the yellow contour of ground-truth in order to generate the green long and short axes. This process simulated a human user manually drawing on MPR. (Color figure online)

The output is a 3D mesh fit to voxel classification by adapting vertices connected by virtual springs to their neighbors to provide a regularizing force that smooths the surface. The true long axis is measured, which may not lie in any orthogonal plane.

2.6   Experiment with Simulated Drawing of Long Axes

As a preliminary experiment, a batch simulation was performed where long and short axes were automatically drawn on each of 285 multispectral MR brain scans of glioma patients in the BraTS 2018 training data. To achieve this, the ground-truth was analyzed to find an appropriate slice on which to draw the long axis. Given the range of slices that contained any ground-truth, the central third of this range was considered, and from that subset, the slice with the largest area of ground-truth was chosen. An automatic process then drew the long axis across the ground-truth on that slice.

In order to simulate the type of long axis that a human user might draw, the axis position was favored to be more medial than the true longest axis. Therefore, on each plane, an ellipse was fit by Principle Component Analysis (PCA) [21] to the segmentation on that slice. The long axis with the same orientation as the major axis of the ellipse was found. The short axis was then found as the longest axis perpendicular to this, as shown in Fig. 6.

The center of the long axis was used for the center of the reformatted sagittal and coronal planes to comprise a 3-plane MPR. Then long and short axes were drawn in similar manner on all planes. The drawn axes precipitate MPR segmentation. Figure 7 shows a few examples.

3   Results

Multi-institutional, routine clinically-acquired pre-operative multispectral MR scans were provided by the 2018 BraTS challenge [24–26]. The data had been preprocessed to be co-registered to the same anatomical template, interpolated to the same resolution (cubic mm), and skull-stripped.

Segmentation accuracy was computed by uploading labeled images to the CBICA Image Processing Portal, which measured statistics for active (enhancing) tumor, whole tumor, and tumor core. While the ground truth was available for the 285 training cases, there were an additional 66 validation cases, and 191 test cases where ground truth was unavailable to participants.

3.1   Experiment with Simulated Drawing of Long Axes

The T1-weighted post-contrast scan was combined with the T2-weighted scan to create a dual-spectra image that was input to the interactive algorithm. Since long axes where drawn on core tumor, this experiment segmented only that structure. Table 3 lists results of the experiment described in Sect. 2.6.
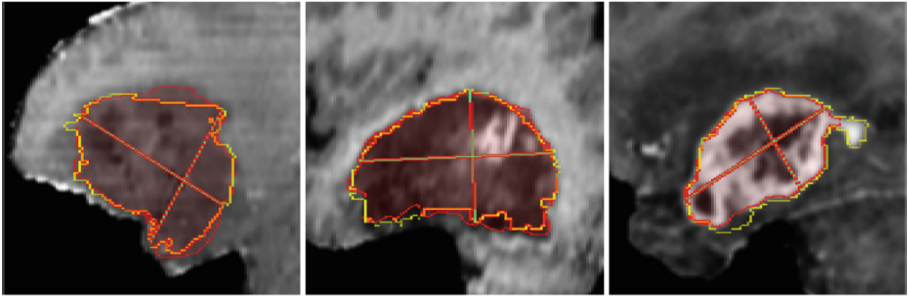
Fig. 7. MPR segmentation (red) depicted relative to ground-truth contours (yellow) and long/short axes (green) on a reformatted sagittal slice. MPR segmentations (not final 3D) were measured to have 0.90 average Dice, compared to ground-truth, for 855 planes of 285 cases. (Color figure online)

Table 3. BraTS 2018 validation results with *simulated* user interaction (core tumor only).

	Dice	Sensitivity	Hausdorff-95%
Mean	0.862	0.893	4.38
Std. Dev	0.060	0.081	3.18
Median	0.873	0.915	3.61

3.2 Experiment with Deep Learning Alone

All four MR modalities of the 66 validation cases were presented to the trained CNN model, and Table 4 presents the results.

Table 4. BraTS 2018 validation results with *no* user interaction.

	Dice			Sensitivity			Hausdorff-95%		
	Active	Whole	Core	Active	Whole	Core	Active	Whole	Core
Mean	0.696	0.878	0.763	0.823	0.893	0.883	6.92	10.1	8.35
Std. Dev	0.200	0.0378	0.256	0.134	0.102	0.127	11.0	16.7	11.5
Median	0.848	0.894	0.918	0.836	0.933	0.938	1.41	3.00	2.83

3.3 Experiment with User Interaction

Ten volunteers used the interface on the BraTS validation and test data. On the 66 scans of the validation data, average Dice coefficient on tumor core improved from 0.76 with deep learning alone, to 0.82 as an integrated, interactive system. The Hausdorff-95% distance improved from 8.4 to 7.5 mm, as detailed in Table 5. Progress since the challenge has improved scores further to 0.87 Dice and 4.9 mm, which is presently the lowest Hausdorff distance on the BraTS leaderboard.

Table 5. BraTS 2018 validation results with *real* user interaction.

	Dice			Sensitivity			Hausdorff-95%		
	Active	Whole	Core	Active	Whole	Core	Active	Whole	Core
Mean	0.730	0.890	0.823	0.752	0.885	0.791	4.37	5.53	7.53
Std. Dev	0.264	0.057	0.173	0.244	0.0865	0.203	6.86	7.06	10.9
Median	0.828	0.904	0.885	0.807	0.904	0.870	2.50	3.87	3.74

On the 191 scans of the test data, the average Dice coefficient was 0.75 and median 0.86. Compared with the simulation experiment where mean and median were quite similar, the disparity between mean and median here suggest that human volunteers and ground-truth disagreed, curtailing certain scores. Table 6 presents the details.

Table 6. BraTS 2018 *test* results with real user interaction.

	Dice			Hausdorff-95%		
	Active	Whole	Core	Active	Whole	Core
Mean	0.643	0.852	0.750	4.88	8.15	8.44
Std. Dev	0.300	0.135	0.267	5.77	12.9	12.7
Median	0.765	0.896	0.864	3.00	4.12	4.24

The Hausdorff-95% distance was nearly a factor of two lower on the simulation experiment, when compared with the other three experiments, which shows the value of knowing precisely where to draw axes. An interesting observation is that the sensitivity for core tumor was higher without user interaction. The median was even slightly higher even though the mean was much lower. This suggests DL fared better than humans on obvious tumors, but users provided essential aid when DL missed badly.

4   Discussion

In comparison with other semi-automatic tools, products from Invivo [27] and Mirada [28] feature initialization by a single click, whereas the additional information contained in SAMBAS' long axis bolsters reliability relative to a click. Perhaps the most similar algorithm to SAMBAS is the GrowCut algorithm [29, 30] implemented in the 3D Slicer [31]. Both have general applicability, and a concept of Background and Foreground regions. However, GrowCut is not initiated as quickly as a drag across the long axis, and one study measured lung lesion contouring to require an average of 10 min [32], whereas a clinical goal is sub-minute. Perhaps the most similar initialization method is [33] for the Random Walker algorithm [34], because a clicked point or stroke commences 2D segmentation from which Background and Foreground seeds are generated for 3D segmentation. However, the SAMBAS approach intentionally

seeks statistical separation rather than a simple circumscribed shape for Background. GrowCut and the Random Walker both lack the two additional regions that SAMBAS adds, Containment and Avoidance, which make editing expeditious. Furthermore, SAMBAS differs by its Bayesian framework, which in conjunction with the added regions, make it possible to seamlessly incorporate organ-specific processing, and to employ DL-based CAD to derive additional SVP probability maps.

Quantitative results were promising, while leaving ample opportunity for improvement. During the interactive experiment, the long axis was drawn manually by human users, with the guidance of real-time MPR segmentation as constructive feedback. The advantage of feedback did not produce better quantitative scores than the batch-generated long and short axes of the simulation experiment. The drop-off in scores between the simulation experiment on training data, and the interactive experiment on validation data, indicates the value of knowing where to draw.

The novel "snap to" feature introduced here may offer a solution to the problem of false positives with CAD systems. Only those CAD findings which are drawn on by the user will be output. The other CAD findings could be withheld from clinicians to avoid biasing their judgment.

The interactive algorithm was developed to be general-purpose, and is well-suited for CT lung and liver lesions. The MR-specific and brain-specific enhancements presented herein are a new addition, which is a work in progress, and we look forward to upgrading the tissue segmentation and deep learning components to improve the overall system. The fact that the integrated system outperformed deep learning alone on the validation data bodes well for interfaces which unite neural networks with expert users.

References

1. Nordstrom, R.: The quantitative imaging network in precision medicine. Tomography **2**(4), 239–241 (2016)
2. Macdonald, D.R., Cascino, T.L., Schold Jr., S.C., Cairncross, J.G.: Response criteria for phase II studies of supratentorial malignant glioma. J. Clin. Oncol. **8**, 1277–1280 (1990)
3. Therasse, P., et al.: New guidelines to evaluate the response to treatment in solid tumors. J. Natl. Cancer Inst. **92**(3), 205–216 (2000)
4. Sorensen, A.G., Batchelor, T.T., Wen, P.Y., Zhang, W.T., Jain, R.K.: Response criteria for glioma. Nat. Clin. Pract. Oncol. **5**, 634–644 (2008)
5. Suzuki, C., et al.: Radiologic measurements of tumor response to treatment: practical approaches and limitations. Radiographics **28**, 329–344 (2008)
6. Wen, P.Y., et al.: Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. **28**, 1963–1972 (2010)
7. Yankeelov, E., Mankoff, D., Schwartz, L., Rubin, D.: Quantitative imaging in cancer clinical trials. Clin. Cancer Res. **22**, 284–290 (2016)
8. https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI_BraTS/MICCAI_BraTS_2017_proceedings_shortPapers.pdf

9. Mehta, A.I., Kanaly, C.W., Friedman, A.H., Bigner, D.D., Sampson, J.H.: Monitoring radiographic brain tumor progression. Toxins **3**(3), 191–200 (2011). https://doi.org/10.3390/toxins3030191

10. Freedman, D.H.: A reality check for IBM's AI ambitions. MIT Technol. Rev. (2017). https://www.technologyreview.com/s/607965/a-reality-check-for-ibms-ai-ambitions/

11. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694

12. Bakas, S., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nat. Sci. Data **4**, 170117 (2017). https://doi.org/10.1038/sdata.2017.117

13. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49

14. Kayahbay, B., Jensen, G., Van Der Smagt, P.: CNN-based segmentation of medical imaging data. arXiv Prepr. arXiv:1701.03056 (2017)

15. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 287–297. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_25

16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

17. Roth, H.R., et al.: An application of cascaded 3D fully convolutional networks for medical image segmentation. Comput. Med. Imaging Graph. **66**, 90–99 (2018)

18. Jackson, P., Hardcastle, N., Dawe, N., Kron, T., Hofman, M.S., Hicks, R.J.: Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy. Front. Oncol. (2018)

19. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 178–190. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_16

20. Multimodal brain tumor segmentation challenge 2017 rankings. https://www.med.upenn.edu/sbia/brats2017/rankings.html

21. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley, New York (2001)

22. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes: The Art of Scientific Computing, 3rd edn. Cambridge University Press, Cambridge (2007)

23. Li, S.Z.: Markov Random Field Modeling in Image Analysis. Advances in Computer Vision and Pattern Recognition. Springer, London (2009). https://doi.org/10.1007/978-1-84800-279-1

24. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. Cancer Imaging Arch. (2017). https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q

25. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. Cancer Imaging Arch. (2017). https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF

26. Bakas, S., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629 (2018)
27. Invivo DynaCAD. http://www.invivocorp.com/solutions/lung-cancer-screening/
28. Mirada XD3. http://www.mirada-medical.com
29. Vezhnevets, V., Konouchine, V.: GrowCut: interactive multi-label ND image segmentation by cellular automata. In: Proceedings of Graphicon, vol. 1, pp. 150–156 (2005)
30. Zhu, L., Kolesov, I., Gao, Y., Kikinis, R., Tannenbaum, A.: An effective interactive medical image segmentation method using fast growcut. In: MICCAI Workshop on Interactive Medical Image Computing, Boston (2014)
31. D Slicer. http://www.slicer.org
32. Velazquez, E.R., et al.: Volumetric CT-based segmentation of NSCLC using 3D-Slicer. Sci. Rep. **3**, 3529 (2013)
33. Jolly, M.P., Grady, L.: 3D general lesion segmentation in CT. In: 5th IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, pp. 796–799. IEEE (2008)
34. Grady, L.: Random walks for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **28**(11), 1768–1783 (2006)