

Confidence-based Ensemble for GBM brain tumor segmentation

Jing Huo¹, Eva M. van Rikxoort¹, Kazunori Okada², Hyun J. Kim¹, Whitney Pope¹, Jonathan Goldin¹, Matthew Brown¹

1 Center for Computer vision and Imaging Biomarkers, Department of Radiological Sciences, David Geffen School of Medicine at UCLA

2 Computer Science Department, San Francisco State University

ABSTRACT

It is a challenging task to automatically segment glioblastoma multiforme (GBM) brain tumors on T1w post-contrast isotropic MR images. A semi-automated system using fuzzy connectedness has recently been developed for computing the tumor volume that reduces the cost of manual annotation. In this study, we propose an ensemble method that combines multiple segmentation results into a final ensemble one. The method is evaluated on a dataset of 20 cases from a multi-center pharmaceutical drug trial and compared to the fuzzy connectedness method. Three individual methods were used in the framework: fuzzy connectedness, GrowCut, and voxel classification. The combination method is a confidence map averaging (CMA) method. The CMA method shows an improved ROC curve compared to the fuzzy connectedness method ($p < 0.001$). The CMA ensemble result is more robust compared to the three individual methods.

1 INTRODUCTION

Glioblastoma multiforme (GBM) is the most aggressive type of primary brain tumor. The median overall survival is 6 months. Clinical trials of new treatments for GBM brain tumors are ongoing. In a phase II trial, radiographic response, in combination with clinical status, is used to assess therapeutic response. Most trials for patients with GBM use the Macdonald criteria, which makes measurement of the maximal enhancing tumor diameter on a single axial image. However, the criteria don't specifically address the presence of necrotic portions of the lesions. Recently, researchers have investigated computer-aided volumetric methods, which can overcome the limitation when tumor contains a non-enhancing necrotic core or has an irregular shape.

GBM brain tumor segmentation is a challenging task due to the variable nature of tumor presentation and the lack of a standard intensity scale in MRI. Brain tumors are heterogeneous; highly variable in size, location, shape and appearance properties; and often deform other nearby structures in the brain. Some artifacts of magnetic resonance (MR) imaging increase the difficulty of tumor segmentation. The imperfection of the radio frequency (RF) pulses and the location of RF coils may introduce non-uniformity of MR images. This study focuses on recurrent GBM brain tumors that develop after surgery, many of which contain a cavity; and the enhanced parts can be a variety of shapes, for example, ring-shape, blob-like shape, or multiple pieces attached to the cavity or spreading into the brain tissue. Furthermore, when the patients are scanned at multiple centers, with different scanners and contrast agent injection, the image intensity contrast varies greatly among different scanners.

Computer-based brain tumor segmentation has remained largely experimental work. Many efforts have exploited MRI's multi-dimensional data capability through multi-spectral analysis⁵⁻⁹. Although multi-channel MR series are useful in differentiating brain tissues and disease, they are usually acquired at low resolutions, with slice gaps and images between different series are often not aligned to each other. High resolution MR volume is often routinely the post contrast T1 weighted image volume. There is little published literature on GBM brain tumor segmentation on a single MR channel. Recently, Liu et al.¹ presented an interactive system using fuzzy connectedness for volume estimation and showed promise in volume estimates on a small number of cases.

In this study, we propose an ensemble approach, applied to semi-automated GBM brain tumor segmentation on T1w post-contrast volumetric MR images, and evaluate the performance on a dataset with 20 cases from a pharmaceutical drug trial. The purpose of the study is to improve the accuracy and robustness of the semi-automated GBM brain tumor segmentation. There are two steps involved. The first is to generate input segmentation candidates from different algorithms; the next is to ensemble them to generate a final result. Three general-purpose segmentation

methods were applied to generate input segmentations: fuzzy connectedness¹, GrowCut³, and voxel classification using a linear discriminant classifier (LDC)⁴. We compare our method to the fuzzy connectedness method.

2 METHODS

The brain volume was preprocessed to remove non-brain matter and obtain consistent image intensities across all subjects for the given MR channel by the following steps: [1]skull-stripping using FSL²; [2]intensity normalization using Freesurfer to standardize the intensity across different scanning centers¹⁰. Then users define a bounding cube as the volume of interest (VOI). The proposed framework consists of two steps: applying different individual segmentation methods and ensemble all segmentation results.

2.1 Individual segmentation methods

We explored three segmentation algorithms as base methods including two semi-automated methods and one learning-based method: GrowCut³, fuzzy connectedness¹, and voxel classification using linear discriminative classifier (LDC)⁴. The GrowCut method was implemented as described in [3] and was picked due to user-friendly interaction, easy implementation and good accuracy in natural image segmentation. To apply the GrowCut method, users needed to click 3 to 5 points on foreground and background structures. The fuzzy connectedness method was implemented as described in Liu et al¹ since it was evaluated for semi-automated GBM brain tumor segmentation. The seed points on the tumor from the GrowCut method were used as the user inputs to the FC method. The voxel classification method was picked as a general-purpose method (LDC) and tuned to this specific application during a learning procedure.

GrowCut and fuzzy connectedness do not require any training and are directly applied to the 20 test scans. In voxel classification, leave-one-tumor-out cross validation was performed. As training data, voxels from inside manually contoured tumors were used as positive examples, an equal amount of voxels sampled outside the tumor were used as background examples. For each training sample, a set of features was calculated: intensity, gradient magnitude, first-order Gaussian derivatives (in three directions), second-order Gaussian derivatives (six in total), and the three eigenvalues of the Hessian matrix on scales 1, 2, and 4, resulting in 42 features in total. A LDC was trained for each leave-one-tumor-out iteration, resulting in 20 runs. To apply the voxel classification to the test scan, the set of 42 features is calculated for each voxel and assigned a probability that it belongs to a tumor by the trained classifier.

2.2 Ensemble all segmentation results into a final result

The three individual methods were combined by confidence map averaging (CMA); the output of the ensemble is the average of the three confidence maps generated by three individual methods. On confidence maps (CM), each individual method shows how confident it is to label each voxel. For LDC method, the output probability map was used as CM. For the GrowCut method, a strength map was generated and used as the CM. For the fuzzy connectedness method, the membership value is used as the CM.

$$CMA(i, j, k) = \frac{1}{N} \sum_{n=1}^N CM(i, j, k)$$

3 EXPERIMENTS AND RESULTS

We used 20 GBM tumors from 18 patients in this study (T1w post-contrast 3D volume with voxel size 0.9*0.9*1mm). The ground truth for the segmentation was manually contoured by a board-certified neuroradiologist.

The performance of the three individual methods and the ensemble method was evaluated using ROC curves (Figure 1). The CMA method shows a significant improvement over fuzzy connectedness in terms of area under the ROC curve ($p < .001$), CMA has the largest area under the curve (AUC), which shows improvement over GrowCut ($p < .001$) but did not reach a significant level over LDC voxel classification ($p > .05$).

The accuracy of the segmentation result was also evaluated by calculating the F-measure (ranging from 0-1) between the semi-automated segmentation result and the ground truth, as shown in Figure 2. To calculate the F-measure, the outputs of the LDC, fuzzy connectedness, and the CMA ensemble were thresholded at an empirically determined threshold of 0.3. With two other methods implemented in addition to fuzzy connectedness, the CMA ensemble performed significantly better than the fuzzy connectedness method ($p < .001$). The ensemble result also showed an

improvement over GrowCut ($p < .05$), but no difference from LDC ($p > .05$). The mean and STD of the F-measure for the 20 test cases is shown in Table 3. The ensemble combination has the lowest STD, which indicates that it is the most robust method. However, it can be seen that not a single individual method beat all of the rest of the individual ones in every tumor case. For all 20 cases, the CMA ensemble result was fairly close to the best method, and more stable among all different tumor cases. As can be seen in Figure 2, on case #1 and #13 where LDC generates poor results, the ensemble still gives as good results as the best individual method. This shows the potential value of the ensemble framework when applied to a larger database.

A visual comparison of the three individual methods and the ensemble method on different slices of one GBM brain tumor case is shown in Figure 3. It can be appreciated that the ensemble result picks up the pieces that were missed by other individual methods.

4 DISCUSSION AND CONCLUSION

In this study, we proposed an ensemble framework for the application of GBM brain tumor segmentation on high-resolution T1w post contrast MR images. The present study introduced the concept of combining the results from different segmentation algorithms for the application to GBM brain tumors. Instead of inventing a customized method for this specific application, the proposed ensemble approach can make use of existing general-purpose segmentation algorithms and combine their results into a more robust and accurate segmentation. GBM brain tumor segmentation on a single channel T1w post-contrast MR scan is a difficult problem. The ensemble framework proposed in this paper improved the performance over a state-of-the-art semi-automated method using fuzzy connectedness, by ensembling with two additional segmentation methods. In addition, the ensemble framework produces the most robust result compared to three individual methods.

There are a couple of limitations in the present study. One of them is that we only had one radiologist's reading as the ground truth (GT). In the future, we will have two board-certified readers' readings as GTs, and compared the performance of the ensemble framework with the inter-reader reproducibility. The other limitation is that the seeds input for FC and GrowCut methods were not strictly controlled among all cases. For future work, we will design a tightly-controlled rule for the user seed inputs to compare the two methods, and vary the number of seeds to study the repeatability.

In conclusion, the ensemble framework using confidence map averaging showed an improvement over fuzzy connectedness and generates more stable and robust results on GBM tumor segmentation. The value of using an ensemble method over individual methods is that it utilizes the complementary strengths of the individual methods and results in more stable performance.

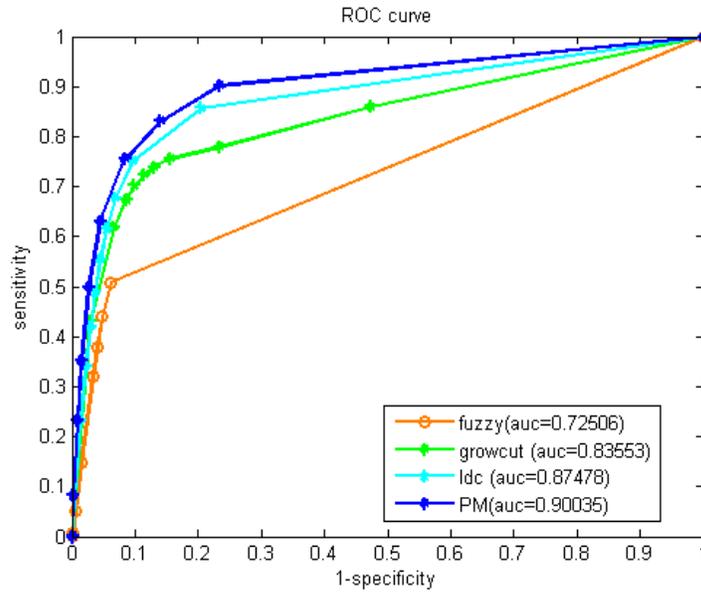


Figure 1: The ROC curve of three individual methods and the CMA ensemble combine three methods

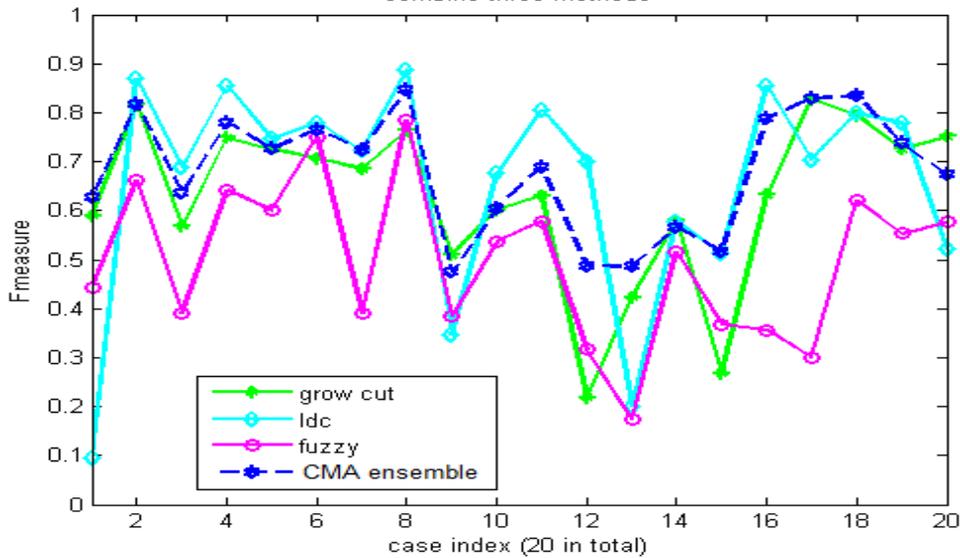


Figure 2: The Fmeasure of the three different methods and the CMA ensemble

Table 1 The mean and STD of the F-measure for 20 cases for different methods

	FC	GC	LDC	CMA
mean \pm STD	0.50 \pm 0.16	0.63 \pm 0.17	0.66 \pm 0.22	0.68 \pm 0.12

REFERENCES

- [1] Liu, J., Udupa, J., Odhner, D., Hackney, D. and Moonis, G., "A system for brain tumor volume estimation via mr imaging and fuzzy connectedness," *Comput. Med. Imaging Graph.* 29(1), 21–34 (2005).
- [2] Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M. and

- Matthews, P.M., "Advances in Functional and Structural MR Image Analysis and Implementation as FSL," *NeuroImage*, 23(S1), 208–219 (2004).
- [3] Vezhnevets, V. and Konouchine V., "Growcut - interactive multi-label n-d image segmentation by cellular automata" *Proc. of Graphicon*, 150–156 (2005).
- [4] Duda, R.O., Hart, P.E. and Stork, D.H., [Pattern classification], Wiley Interscience, MA, USA, (2000).
- [5] Phillips, W.E., Velthuizen, R.P., Phuphanich, S., Hall, L.O., Clarke, L.P. and Silbiger, M.L., "Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme," *Magn Reson Imaging* 13(2), 277-290 (1995).
- [6] Clark, M., Hall, L., Goldgof, D., Velthuizen, R., Silbiger, M. and Murtagh, F., "Automatic tumor segmentation using knowledge-based techniques," *IEEE Trans Med Imag.* 17(2), 187-201 (1998).
- [7] Ho, S., Bullitt, E., and Gerig, G., "Level-set evolution with region competition: automatic 3-D segmentation of brain tumors," *Pattern Recognition, Proceedings. 16th International Conference on* , 1, 532-535 (2002).
- [8] Corso, J. J., Sharon, E., Dube, S., El-Saden, S., Sinha, U. and Yuille, A.L., "Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification," *IEEE Transactions on Medical Imaging*, 27(5), 629-640 (2008).
- [9] Prastawa, M., Bullitt, E., Ho, S. and Gerig, G., "A brain tumor segmentation framework based on outlier detection." *Medical Image Analysis*, 8(3), 275-283 (2004).
- [10] Zijdenbos, A.P. Evans A.C. and Sled, J.G., "A nonparametric method for automatic correction of intensity nonuniformity in mri data." *IEEE Trans Med Imaging*, 17(1), 87-97 (1998).

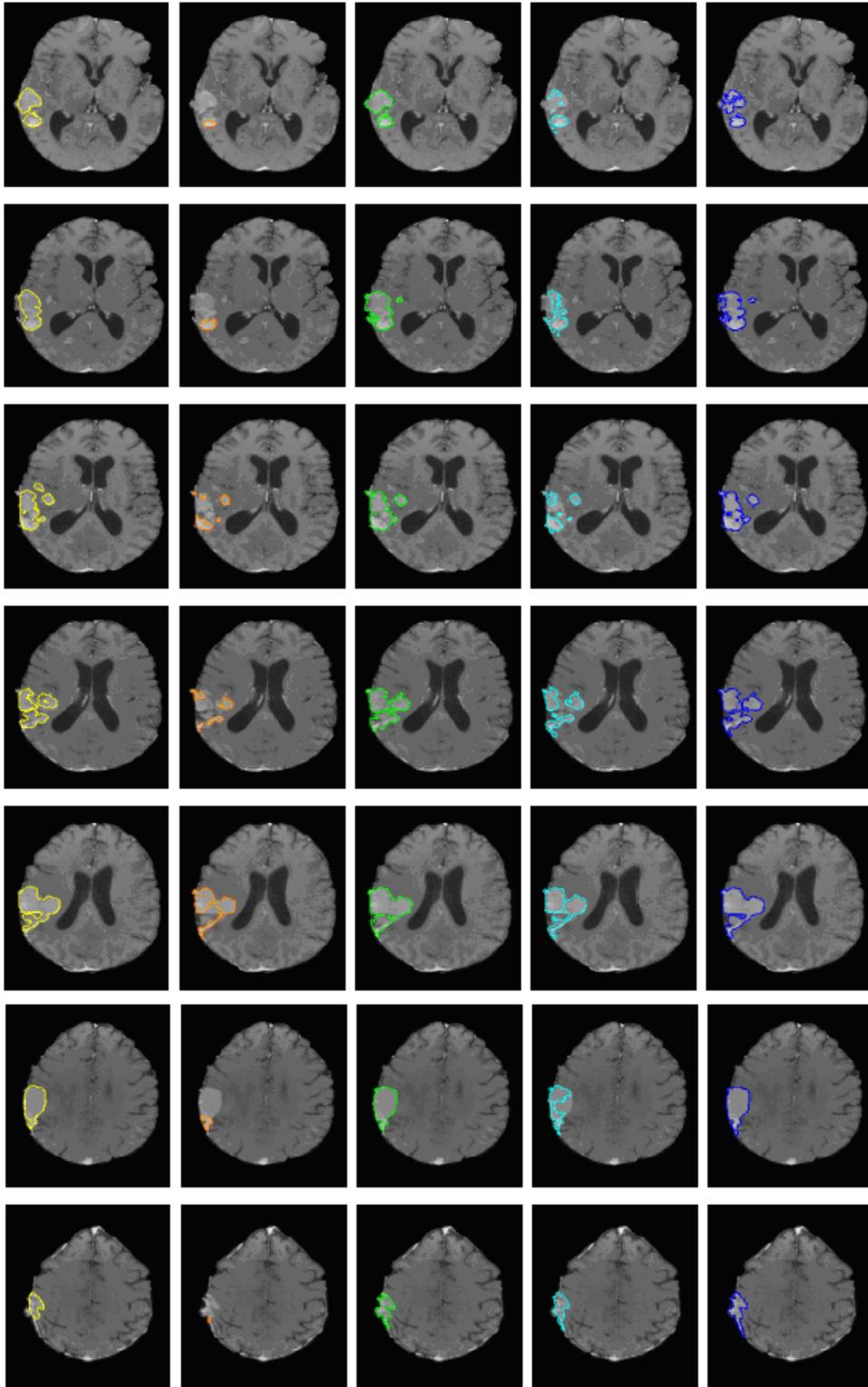


Figure 3. Illustrative example of one GBM tumor. Rows show different slices; columns show different segmentation methods: the ground truth; fuzzy connectedness; GrowCut; LDC; CMA.