Improving semi-automated segmentation by integrating learning with active sampling

Jing Huo[1], Kazunori Okada[2], Matthew Brown[1]
1 Department of Radiological Sciences, David Geffen School of Medicine, University of California in Los Angeles
2 Department of Computer Science, San Francisco State University

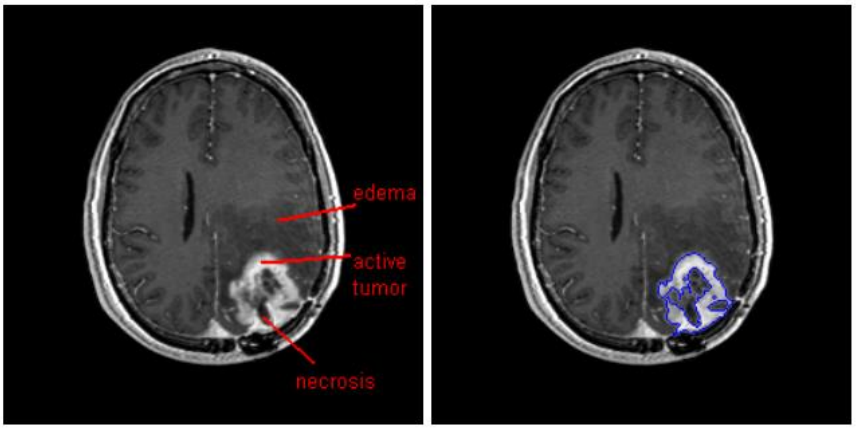
Abstract

Interactive segmentation algorithms such as GrowCut usually require quite a few user interactions to perform well, and have poor repeatability. In this study, we developed a novel technique to boost the performance of the interactive segmentation method GrowCut involving: 1) a novel "focused sampling" approach for supervised learning, as opposed to conventional random sampling; 2) boosting GrowCut using the machine learned results. We applied the proposed technique to the glioblastoma multiforme (GBM) brain tumor segmentation, and evaluated on a dataset of ten cases from a multiple center pharmaceutical drug trial. The results showed that the proposed system has the potential to reduce user interaction while maintaining similar segmentation accuracy.

1. Introduction

Two weaknesses in the interactive segmentation algorithm include the need of excessive user interactions and the lack of repeatability. The lack of repeatability is due to the interaction with the segmentation process. The excessive user interaction means that many of the interactive segmentation methods need a large number of input seeds to perform well. In this study, we performed GrowCut[1] as the interactive method for glioblastoma multiforme (GBM) brain tumor segmentation on T1w post-contrast images in clinical trials.

GBM brain tumor segmentation is the most severe brain tumor type. It usually consists of three parts: active tumor, necrosis and edema, as shown in Figure 1. In this study, we segmented the active tumor part, that is, contrast-enhancing component on MRI, as defined in RANO criteria[2] for treatment response evaluation, for measurable target lesions.



a) b)

Figure 1 GBM brain tumor: a) the tumor presentation on a T1w post contrast MR image slice; b) the tumor contour drawn by a neuroradiologist

GBM brain tumor segmentation on a single channel of T1w post contrast MR imaging is a challenging task due to the large variation in appearance, shape and locations. Another difficulty is that the dataset is collected from multiple medical centers in phase II trials. Images acquired from different centers and scanners by nature have different intensity range, noise level, etc, and thus vary in the presentation of the feature space.

There are several problems when applying GrowCut to GBM tumor segmentation. First, the necrosis could not be picked up with minimal seed on active tumor parts. Second, for multi-focal tumors where there are unconnected tumor pieces, GrowCut cannot detect all of them given minimal seeding on one tumor component. Third, on post-contrast MR images, vessels and dura appear bright, as well as active tumors. GrowCut cannot differentiate between active tumors and other bright non-tumor structures as shown on Figure 5. In summary, for original GrowCut method, readers usually need to click quite a few seeds to update the GrowCut segmentation, in order to generate good results.

Researchers have studied supervised learning to perform fully-automated GBM tumor detection and segmentation[3]. Conventionally, voxels from inside manually-contoured tumors are used as positive examples and a similar number of voxels outside tumors were used as background examples. The background samples include white matter, gray matter, ventricle, vessels, necrosis, edema, dura, etc. The weakness of random sampling in classifier learning is that the bright structures of vessels and dura cannot be correctly segmented due to the deficiency of training samples in feature space.

In this study, we developed "focused sampling" to only sample the "difficult" voxels, to improve upon conventionally random sampling. In the training phase, specific samples are selected that are "difficult" for the GrowCut method. In the testing phase, we boosted GrowCut using learning results from focused sampling to help reduce the user inputs and to improve the repeatability. The hypothesis is that by integrating machine learning with focused sampling, the boosted GrowCut will reduce the user interaction while maintaining similar accuracy, and improve the repeatability. There is little work done improving interactive segmentation with learning[4]. To our knowledge, ours is the first to employ "focused sampling" to help with interactive segmentation for medical images.

2. Materials and Methods

The brain volumes were pre-processed by skull-stripping using FSL tools[5] and intensity normalization with Freesurfer[6].

2.1 GrowCut

The GrowCut method uses the cellular automata theory. The method starts from user-clicked seeds on both object and background. For each voxels, it allows all the neighbor voxels to attack, and the strength of attacking is based on the local neighbor similarity. The method is run multiple iterations, till the label map and strength map don't change any more. Eventually the method assigns each voxel with both a label and a strength value. The label shows which category it belongs, while the strength shows how confident it is about the labeling.

2.2 Supervised learning with focused sampling

"Focused sampling" means that we collect more training samples for difficult structures and less training samples for easy structures, compared to random sampling where all non-tumor structures are equally treated. Specifically, the "difficult structures" are those that GrowCut method assigned the wrong labels to.

The reason of using "focused samples" for training is that they are the most representative samples of the "difficult structures" for GrowCut. Instead of exploring the whole brain structures, we focus on the structures which are most difficult for GrowCut method. Therefore, the proposed classification system will not do a simple "tumor" vs "non-tumor" classification, instead, it is the "tumor" vs "difficult non-tumor" classification. The "difficult" here means that it is hard to differentiate for GrowCut method.

Our motivation for "focused sampling" actually came from active learning. In active learning, the machine learner automatically picks the difficult samples and asks the oracle for annotation; while in our "focused sampling", readers inherently collect difficult samples for the machine learner, while using interactive segmentation method to be boosted.

The training sample collection is done by readers when using the semi-automated interactive GrowCut method to update the segmentation results. Readers need to follow the three steps: 1) User clicks minimum amount of seeds; 2) Run GrowCut method; 3) Review the segmentation result, gradually paint more seeds on the false-segmented structures, re-run GrowCut and update the result; 4) Repeat step 3) until the user has nothing to modify; 5) In the end, all the seeds user clicked and painted will be saved for future training.

In the training phase, equal number of voxels (n=2 in this study) was randomly sampled from each stroke of reader's painting as training data. 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 linear discriminate classifier (LDC)[7] was trained for each leave-one-tumor-out iteration, resulting in six runs. We used the LDC classifier implementation in prtools[7].

2.3 Boosting GrowCut with learning results from focused sampling

2.3.1 Boosting GrowCut Algorithm

To apply the proposed system to the test scan, there are two phases: first, run the LDC classifier on the current test case; second, reader click initial seeds and run Boosting GrowCut method as described in Algorithm 1. We described our algorithm using the same format as in the original GrowCut paper for comparison, and our proposed changes were highlighted in red.

Algorithm 1: Boosting GrowCut algorithm, with learning result from focused sampling

Each voxel p is represented as a vector $\langle l_p, \theta_p, C_p, P_p \rangle$ - the current label of the voxel, the "strength" of labeling, the intensity, and the LDC classifier output probability. $N(p)$ is the 26-neighborhood system. At iteration $t+1$, voxel label l_p^{t+1} and strength θ_p^{t+1} are updated as follows:

```
// For each cell …
1 For ∀p ∈ P
2     // copy previous state
3     l_p^{t+1} = l_p^t;  θ_p^{t+1} = θ_p^t;
4     // neighbors try to attack current cell
5     For ∀q ∈ N(p)
6         If [w * g(‖C_p − C_q‖) + (1 − w) * g(‖P_p − P_q‖)] * θ_q^t > θ_p^t
7             l_p^{t+1} = l_q^t;
8             θ_p^t = [w * g(‖C_p − C_q‖) + (1 − w) * g(‖P_p − P_q‖)] * θ_q^t;
9         End if
10    End for
11 End for
12
13 Run 1-11 till convergence
14 Flip the label l_p, when P_p > 0.9 & l_p == 0 or P_p < 0.1 & l_p == 1
15 Run 1-11 till convergence
```

Where g(x) is a monotonically decreasing function bounded to [0,1]:
$g(x) = 1 - \frac{x}{max\{\|x\|\}}$

2.3.2 Parameter selection using ROC analysis

In line 8 of Algorithm 1, the strength update is the weighted sum of intensity similarity and LDC posterior probability similarity. The idea is to combine original GrowCut result and LDC classifier result. The parameter w, the weight assigned to original GrowCut, is determined by ROC analysis.

We vary the value of w from 0 to 1 with step 0.1. For each fixed w, we obtain the weighted sum of GrowCut segmentation strength map and the LDC posterior probability P_p^{final}: $P_p^{final} = w * \theta_p + (1 - w) * P_p$. The ROC curve for one tumor case was achieved by varying the threshold on P_p^{final} into binary segmentations and the area under the curve (AUC) is calculated, and five tumor cases were utilized to obtain an averaged AUC for each fixed w as shown in Figure 2. In the end, w is determined to be the one with the greatest AUC with the save five cases.

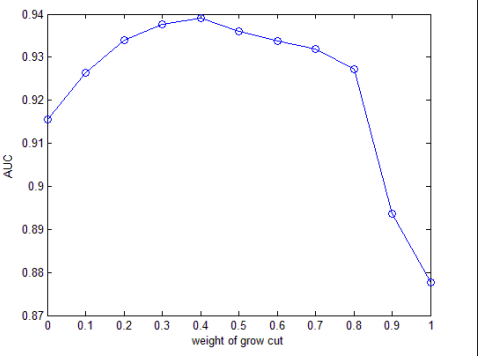


Figure 2 The area under the curve (AUC) with varying the parameter w.

3 Experiments and Results

We have ten GBM tumor cases from three medical centers in this study (T1 weighted post-contrast 3D volume with voxel size 0.9*0.9*1mm and in-plane resolution 256*256). The ground truth for the segmentation was manually contoured by a board-certified neuroradiologist.

We simulated the user interaction seeds and compared the performances between the two systems: original GrowCut, and boosting GrowCut. Both systems were run within a pre-defined volume of interest (VOI). The accuracy of the segmentation was evaluated by calculating the overlap ratio between the ground truth and the segmentation results.

The simulating seeds were randomly picked from tumor and non-tumor regions, excluding the "band region" which is defined as the region between the 3-voxel erosion and dilation of the tumor contour, in order to avoid partial volumed voxels. The segmentation is initialized by one tumor seed and one non-tumor seed. Original GrowCut and boosting GrowCut were applied respectively using the same initialization. After convergence, one additional seed for either tumor or non-tumor seed was added, the seed was randomly picked according to the difference between the result of segmentation and the ground truth excluding the "band region". In our experiment, additional seeds were contiguously added in the same fashion for ten iterations, and the accuracy was plotted as in Figure 4, where y-axis is the overlap ratio, and x-axis is the number of iterations; each iteration one tumor seed and one background seed is added. The initial seeding was totally random, and we ran 20 passes to generate the error bars on the plot. Examples of segmentation results of cases 3, 4, 5 and 9 are illustrated in Figure 5. Figure 3 showed the posterior output of LDC classifier results, where vessels are successfully classified from focused sampling.

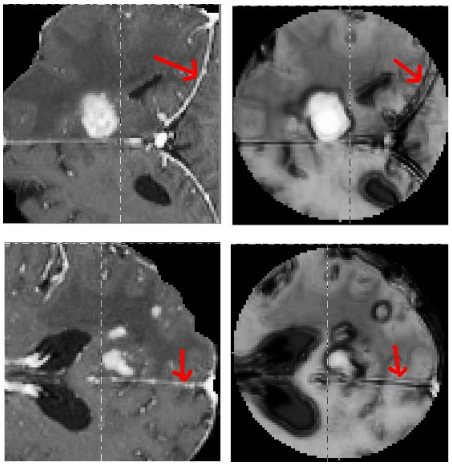
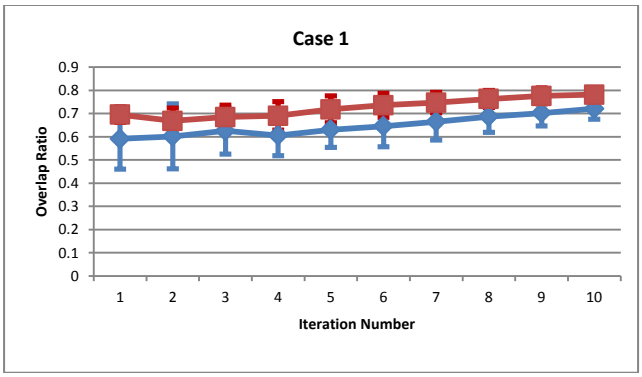
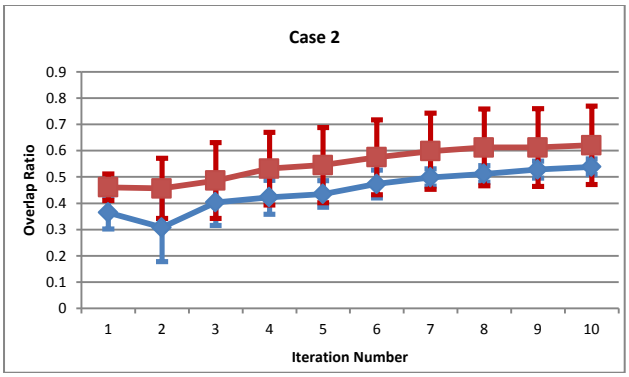
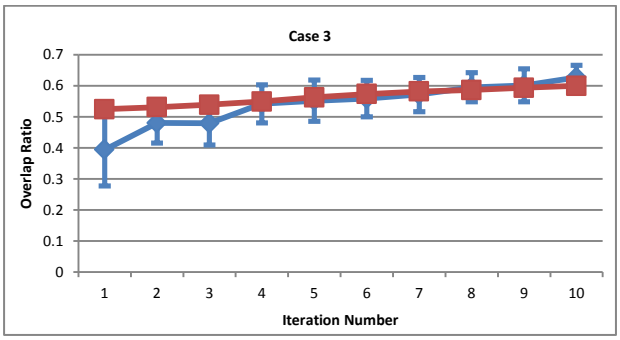
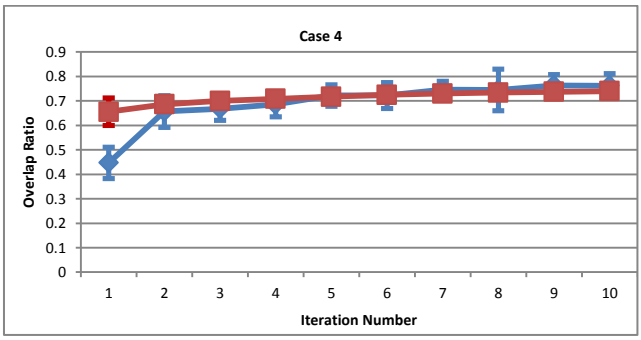


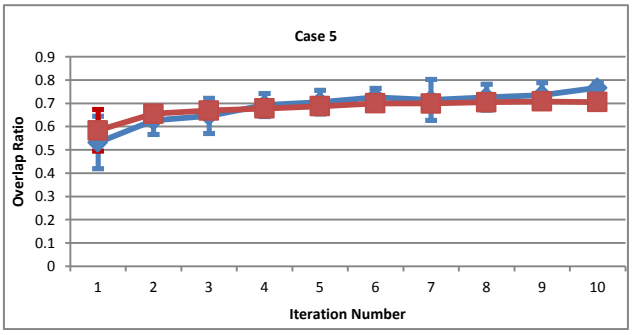
Figure 3 Learning results with focused sampling: two rows are two examples of tumors; two columns are original MR image slice and the classifier probability output respectively. Red arrows show the vessels.

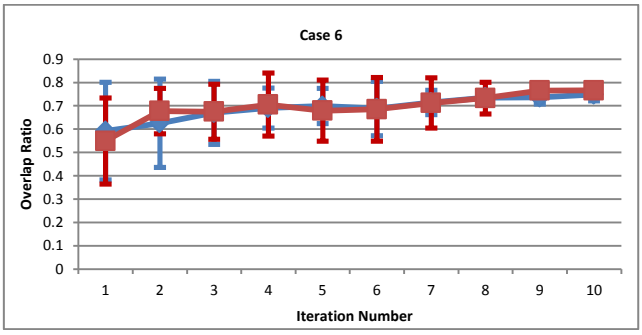

Case 1
Overlap Ratio
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
1 2 3 4 5 6 7 8 9 10
Iteration Number

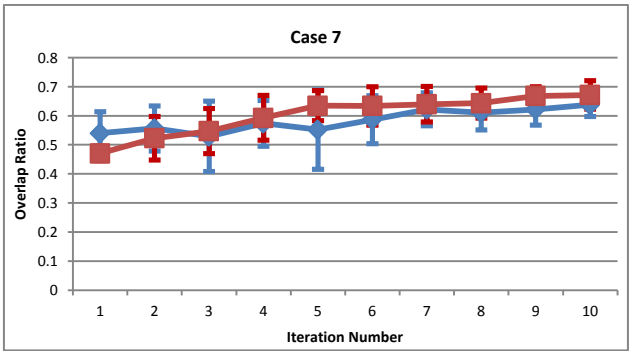

Case 2
Overlap Ratio
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
1 2 3 4 5 6 7 8 9 10
Iteration Number

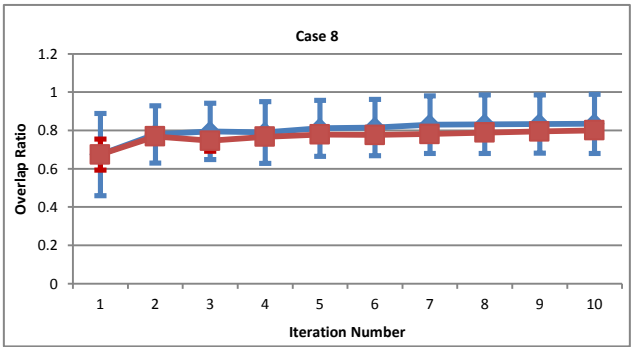

Case 3
Overlap Ratio
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
1 2 3 4 5 6 7 8 9 10
Iteration Number


Case 4
Overlap Ratio
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
1 2 3 4 5 6 7 8 9 10
Iteration Number


Case 5
Overlap Ratio
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
1 2 3 4 5 6 7 8 9 10
Iteration Number


Case 6
Overlap Ratio
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
1 2 3 4 5 6 7 8 9 10
Iteration Number


Case 7
Overlap Ratio
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
1 2 3 4 5 6 7 8 9 10
Iteration Number


Case 8
Overlap Ratio
0 0.2 0.4 0.6 0.8 1 1.2
1 2 3 4 5 6 7 8 9 10
Iteration Number

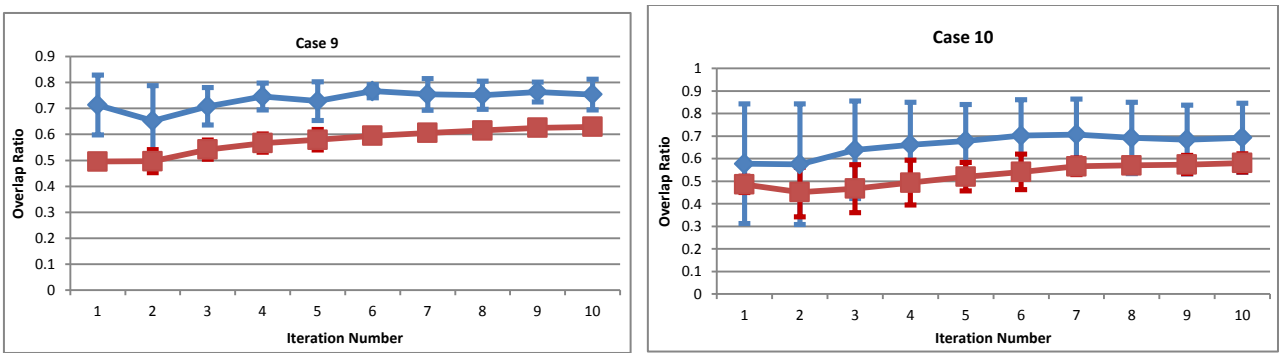


Figure 4 Comparing the original GrowCut (blue dots) and the boosting GrowCut (red squares): at each iteration, one tumor and one background seed is added to improve the segmentation result. The error bars are generated by 20 runs of random initialization.

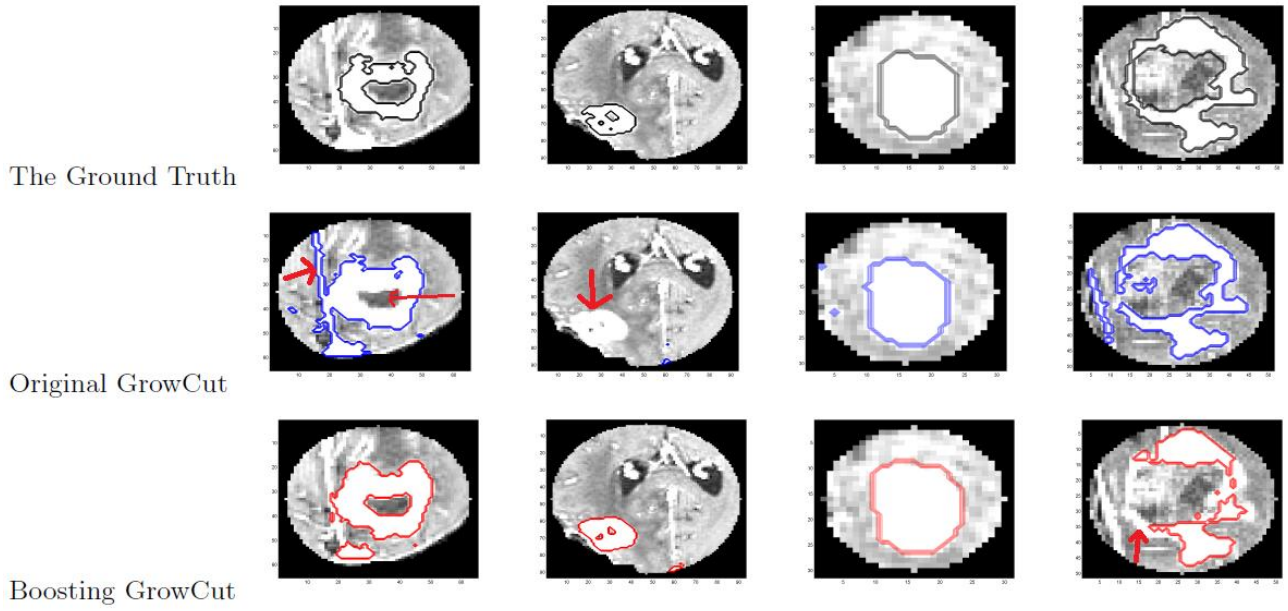


Figure 5 Examples of segmentation results: columns are Case 3,4,5 and 9; rows are the ground truth (black contour), the original GrowCut results (blue contour) and the boosting GrowCut results (red contour)

4 Discussion

We observed that cases 1-4 showed that boosting GC improved the accuracy upon original GrowCut. The reason is that these cases either included necrosis, vessels or an additional unconnected tumor piece. Those structures cannot be correctly segmented by original GrowCut, whereas they were correctly segmented by boosting GrowCut. As shown in Case 3(the first column) of Figure 5, the original GrowCut mistakenly segmented the connected vessel and did not remove necrosis, while boosting GrowCut successfully overcame both problems. Take vessels for example, Figure 3

showed the posterior output of LDC learning results, where vessels are successfully classified from focused sampling. In Case 4 (the second column) of Figure 5, the tumor piece shown was a 2nd unconnected component and was completely missed by original GrowCut, whereas it was correctly segmented by boosting GrowCut.

In cases 9-10, boosting GrowCut performance was worse than original GrowCut. The reason is that both tumors contain a "ring shape" pattern, which shows very thin contrast enhancement around necrosis or surgical cavity, as shown in Case 9 (the fourth column) of Figure 5. There were no samples of this "ring" selected during training and it was thus misclassified by LDC classifier and the segmentation result of boosting GrowCut was deteriorated.

For cases 5-8, boosting GrowCut did not show obvious improvement, there are two reasons. One is that the image contains an isolated and contiguous tumor which does not include any of the structures mentioned above and original GrowCut generates good results as shown in Case 5 (the third column) of Figure 5. The second reason is that the tumor contains both vessels and ring-shaped pattern where the benefits and disadvantages of boosting GrowCut are canceled.

In summary, the boosting GrowCut with focused sampling shows potential in improving the performance. It still has limitations in terms of gathering an adequate range of samples during learning, and we will apply iterative learning in the future. The system of improving interactive segmentation using focused sampling is feasible in a challenging medical application.

References

[1] Vezhnevets, V., Konouchine, V., "Growcut-interactive multi-label n-d image segmentation by cellular automata," Proceedings of Graphicon, 150-156(2005).
[2] Wen, P.Y., Macdonald, D.R., Reardon, D.A., et al., "Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group," J Clin Oncol, 28, 1963–72(2010).
[3] Bauer, S., Nolte, Lutz-P., Reyes, M., "Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization," Proceeding of Medical Image Computing and Computer Assisted Intervention, 6893, 354-361 (2011).
[4] Veeraraghavan, H., Miller, J.V., "Active learning guided interactions for consistent image segmentation with reduced user interactions," Biomedical Imaging: From Nano to Macro, IEEE International Symposium on , 1645-1648 (2011).
[5] Smith, S.M., Jenkinson, M., Woolrich, M.W. et al., "Advances in Functional and Structural MR Image Analysis and Implementation as FSL," NeuroImage, 23(S1), 208-219(2004).
[6] Sled, J.G., Zijdenbos, A.P., Evans, A.C., "A non-parametric method for automatic correction of intensity non-uniformity in MRI data," IEEE Trans. Med. Imag., 17, 87–97(1998).
[7] Duda,R.O., Hart, P.E., Stork, D.H., [Pattern classification],Wiley Interscience, MA, USA(2000).
[8] Duin, R.P.W., "PRTools - Version 3.0 - A Matlab Toolbox for Pattern Recognition," Proc. of SPIE, 1331(2000).