

Anatomy-Guided Brain Tumor Segmentation and Classification

Bi Song^(✉), Chen-Rui Chou, Xiaojing Chen, Albert Huang,
and Ming-Chang Liu

US Research Center, Sony Electronics Inc., San Jose, USA
bi.song@am.sony.com

Abstract. In this paper, we consider the problem of fully automatic brain tumor segmentation in multimodal magnetic resonance images. In contrast to applying classification on entire volume data, which requires heavy load of both computation and memory, we propose a two-stage approach. We first normalize image intensity and segment the whole tumor by utilizing the anatomy structure information. By dilating the initial segmented tumor as the region of interest (ROI), we then employ the random forest classifier on the voxels, which lie in the ROI, for multi-class tumor segmentation. Followed by a novel pathology-guided refinement, some mislabels of random forest can be corrected. We report promising results obtained using BraTS 2015 training dataset.

1 Introduction

Segmentation of brain tumor from medical images is of high interest in surgical planning, treatment monitoring and is gaining popularity with the advance of image guided surgery. The goal of segmentation is to delineate different tumor structures, such as active tumorous core, necrosis and edema. Typically, this process requires several hours of a clinician's time to manually contour the tumor structures. As manual processing is so labor intensive, automated approaches are being sought. Automatic brain tumor segmentation is challenging due to the large variation in appearance and shape across patients.

Most state-of-the-art methods sample the entire MRI volume data to build classifier for multi-class tumor segmentation, which involve high demand of computation and memory. In this paper, we propose a two-stage automatic segmentation method. We first segment the whole tumor by utilizing anatomy structure information for data intensity normalization and tumor separation from non-tumor tissues. Using this initially segmented tumor as a ROI, we then employ a random forest classifier followed by a novel pathology based refinement to distinguish between different tumor structures. As we only apply classification on the voxels within a ROI, our algorithm is more efficient in terms of both time and memory. The workflow of proposed method is shown in Fig. 1.

We provide an empirical evaluation of our method on publicly available BraTS [11] 2015/2016 training set, and compare with the top performing algorithms. The results demonstrate that the proposed method performs the best

in segmentation of active tumor core, and comparably to the top performing algorithms in the other tumor structures.

2 Method

In this section, we present the technical details of our proposed methods, as shown in Fig. 1, including data normalization, initial segmentation, feature extraction, voxel classification and refinement.

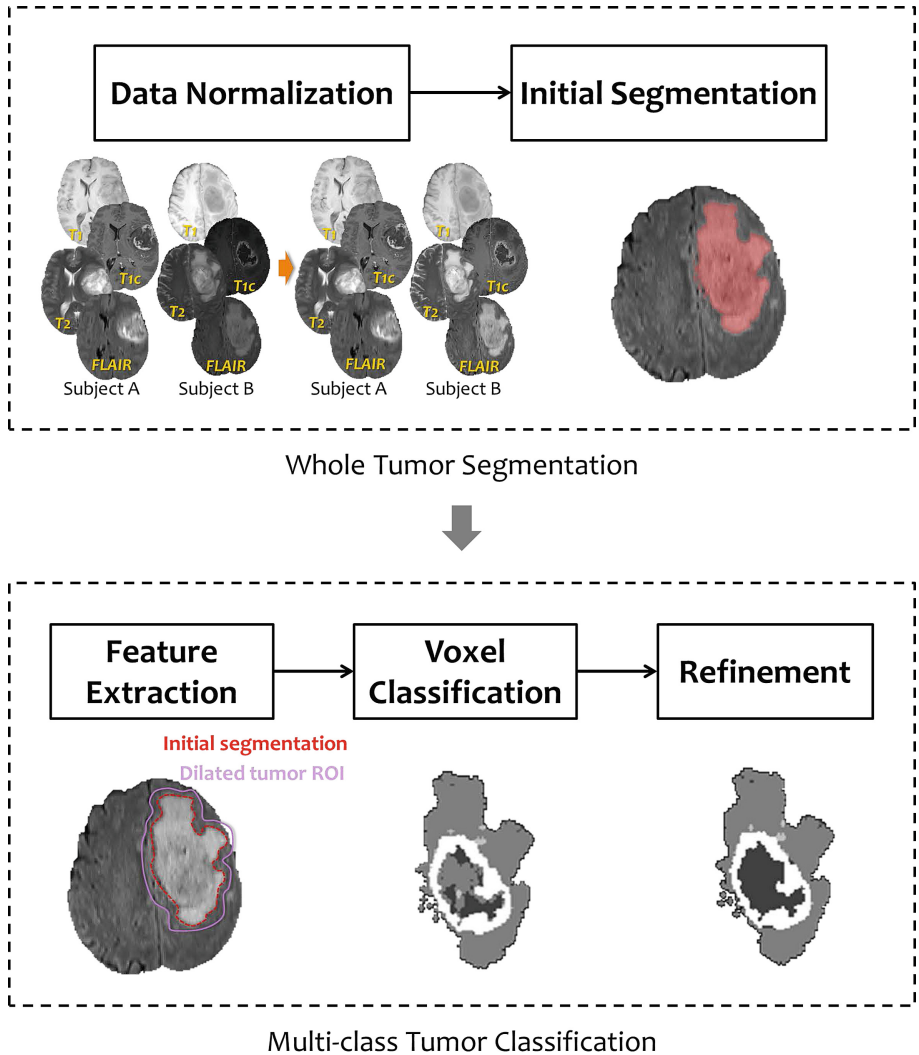


Fig. 1. The proposed two-stage brain tumor segmentation workflow.

2.1 Data Normalization

Intensity inhomogeneities appearing in MRI produce spatial intensity variations of the same tissue over the image domain. To correct the bias field, we applied the N4ITK approach [12]. However, there are large intensity variations across brain tumor MRI data sets and intensity ranges are not standardized; bias correction is not enough to ensure that the intensity of a tissue type across different subjects or even different scans of same subject lie in a similar scale. In [9], a cerebrospinal fluid (CSF) normalization technique is proposed to normalize each individual modality with the mean value of the CSF. However, just utilizing CSF information is not enough, the intensities of other structures need to be aligned as well. To normalize the intensity of imaging data, we propose an anatomy-structure-based method based on the assumption that the same structures of same modality (T1, T1c, T2, Flair), such as white matter (WM), grey matter (GM) or cerebrospinal fluid (CSF), should have similar intensity value across different data sets. To be specific, we apply fuzzy C-means algorithm (FCM) [2, 5] to classify the input data into WM, GM and CSF. Then the normalization is performed by aligning the median values of these WM, GM, and CSF classes for each modality and do piecewise linear normalization in between. Thus ensure these tissue types have similar intensity across the image datasets of same modality. Figure 2 shows an example of intensity histograms before and after data intensity normalization.

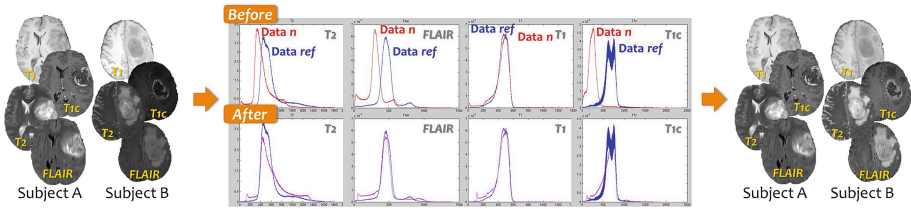


Fig. 2. An example of data normalization results. Note that contrary to just applying histogram matching, the proposed data normalization scheme provides similar intensity scale across subjects and meanwhile keeps the specific structure of each subject’s data.

2.2 Initial Segmentation

Among different modalities of MRI, Flair and T2 provide better boundary contrast of the whole tumor. In [13], symmetric template difference is used as a feature in their supervised segmentation framework. In our method, we also explore symmetric information. By assuming tumor rarely happens completely symmetrically, symmetric differences of Flair and T2 are calculated for locating the initial seeds of tumors. After thresholding and union of the symmetric differences of Flair and T2, we remove the connected components whose size is too small, to reduce the false positive introduced by noise. By selecting the center of

the initial seeds as the target seeds and a bounding box as the background seeds, the GrowCut algorithm [14] is employed on the linear combination of Flair and T2, $\alpha \cdot I_{Flair} + (1 - \alpha) \cdot I_{T2}$, for segmenting the whole tumor. An illustration of initial segmentation is shown in Fig. 3.

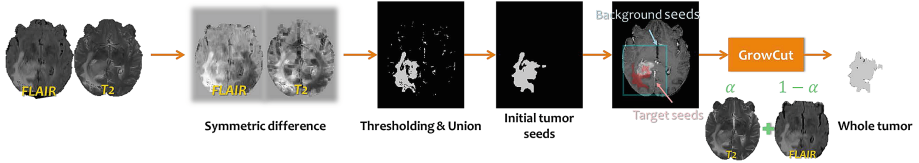


Fig. 3. An illustration of initial tumor segmentation.

2.3 Feature Extraction

The initial segmentation results are dilated to provide a ROI for further multi-class segmentation. The features are extracted from the ROI. Our features include voxel-wise and context-wise features. The voxel-wise features are composed of appearance features, texture features and location features. The context-wise features [7] aim to capture the neighborhood information.

- *Appearance*: Voxel’s intensity value of smoothed T1, T1c, T2 and FLAIR. Gaussian kernel is applied to suppress the data noise.
- *Texture*: Variance of T2 and Laplacian of Gaussian (LoG) on T2, which represent local inhomogeneity.
- *Location*: Initial segmentation results indicating the prior information about the location of tumor.
- *Context*: Multiscale local mean intensity within a box of different size centered on each voxel to catch neighborhood information. Context features are combined from T1c and T2.

An illustration of extracted features is shown in Fig. 4.

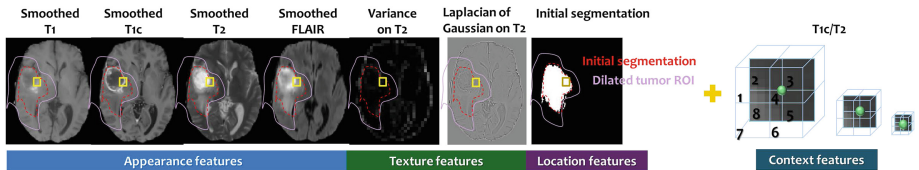


Fig. 4. An illustration of extracted features. The voxel-wise features are shown on the left, and the context-wise features are shown on the right.

2.4 Voxel Classification

A random forest classifier [3] is used for multi-class classification of pixels into five classes: (i) label 1 for necrosis (ii) label 2 for edema (iii) label 3 for non-enhancing tumor, (iv) label 4 for enhancing tumor and (v) label 0 for all other tissues. As illustrated in Fig. 5, each tree outputs a probability of tumor class. The final label of each voxel is decided based on the majority voting of the probabilities.

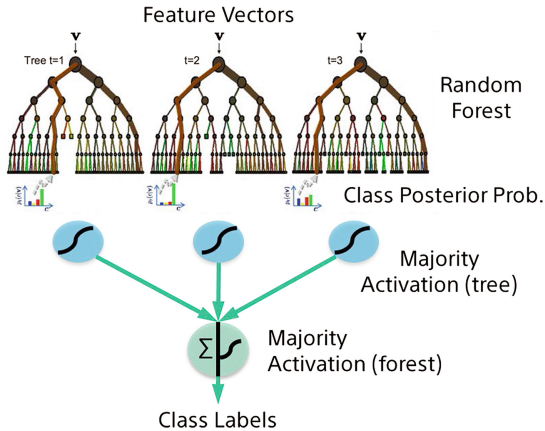


Fig. 5. An illustration of voxel classification using random forest.

2.5 Refinement

Pixel misclassification error might occur in the random forest classification results due to overlapping intensity ranges. For example, necrosis and non-enhancing cores may be mislabeled as edema. We propose a pathology-guided refinement scheme to correct the mislabels based on pathology rules, such as edema is usually not inside the active cores, and non-enhancing cores often surround active cores. Figure 6 shows example results before and after refinement. In Fig. 6(a), the output of random forest classification is shown in the middle, the necrosis inside the active cores is incorrectly labeled as edema. The results after refinement are shown on the right, these errors are corrected based on the pathology-guided rules. In Fig. 6(b), middle shows the output of random forest classification. The non-enhancing core is mislabeled as edema. By identifying core/non-core seeds from the random forest results, these errors can be corrected as shown on the right.

3 Experimental Results

To evaluate the performance of our method, we show results on the BraTS 2015 training data set (identical to BraTS 2016 training data set) [11], which contains

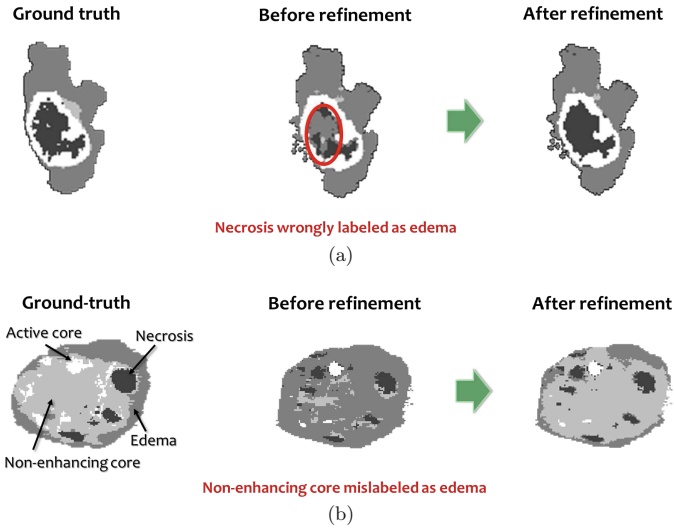


Fig. 6. Example of results before and after refinement. (a) Left: Ground truth, Middle: Random forest output, necrosis inside the active cores is wrongly labeled as edema. Right: Results after refinement. (b) Left: Ground truth, Middle: Random forest output, non-enhancing core is mislabeled as edema. Right: Results after refinement.

220 high-grade and 54 low-grade glioma cases. The dataset include MRI with four different sequences: T1, T1 after gadolinium enhancement (i.e., T1c), T2 and FLAIR. The data volumes are already skull-stripped and registered intra-patient. The volumes include four tumor labels: necrosis, edema, non-enhancing core and tumor active core. The result of our classification framework is a label at every voxel in the 3D MRI volumes. The accuracy measures employed are Dice’s coefficient and Hausdorff distance. Similar to the Virtual Skeleton Database (VSD) [1] online evaluation system for BraTS, the metrics are evaluated on three structures: the “whole” tumor (i.e., all four tumor structures), the tumor “core” (i.e., all tumor structures except “edema”) and the “active” tumor (i.e., only the tumor active core). We perform a leave-one-out cross validation. Note that we do not take high-grade or low-grade as a prior knowledge during training and testing.

3.1 Qualitative Results

Examples of our tumor classification results are shown in Fig. 7. The results are shown on one high grade tumor case and one low grade tumor case along with the corresponding T1, T1c, T2 and FLAIR slices. In both cases, it can be seen that visually our results are comparable to the ground truth labeling. The Dice scores (%) for these two cases are: Case 1 (HGG) Necrosis 85.22, edema 91.45, non-enhancing core 9.19 and tumor active core 94.79; Case 2 (LGG) Necrosis 80.17, edema 66.62, non-enhancing core 56.31 and tumor active core 65.90.

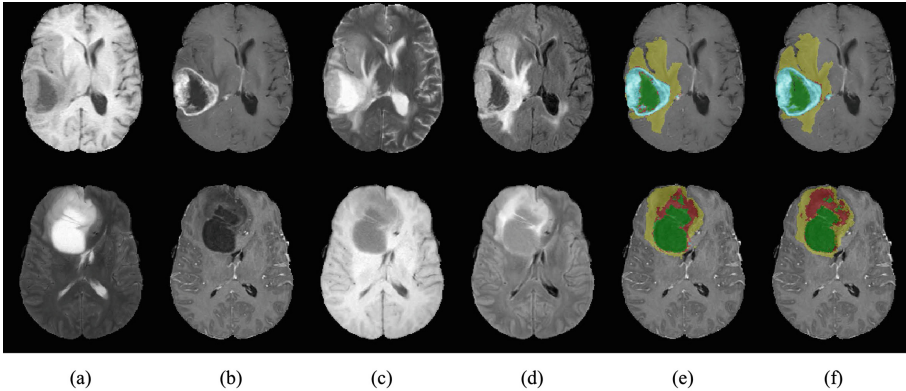


Fig. 7. Top row: high grade tumor case. Bottom row: low grade tumor case. (a) T1 slice, (b) T1c slice, (c) T2 slice, (d) FLAIR slice, (e) ground truth labeling and (f) labels produced by our algorithm (Necrosis in green, edema in yellow, non-enhancing core in red and tumor active core in cyan). (Color figure online)

3.2 Quantitative Results

Table 1 shows the average Dice and Hausdorff distance obtained using our method on a total of 274 cases. The boxplots of Dice and Hausdorff distance are shown in Fig. 8.

Table 1. Results obtained on BraTS 2015 Training dataset, reporting average Dice coefficient and Hausdorff distance. Dice scores for active tumor are calculated for high-grade cases only.

	Whole	Core	Active
Dice (%)	87.0	72.2	75.4
Hausdorff distance (mm)	9.3	9.1	6.5

We compare our results on BraTS 2016 training data set with top performing algorithms in testing phase of the BraTS 2016 Challenge in Table 2. The results show that the proposed method performs the best in segmentation of active tumor core, and comparably to the top performing algorithms in the other tumor structures.

Run time. The run time of training using 254 cases on a computer with 2.67 G Hz GPU, 24 G memory is about 15 min and run time of testing one case is about 1 min.

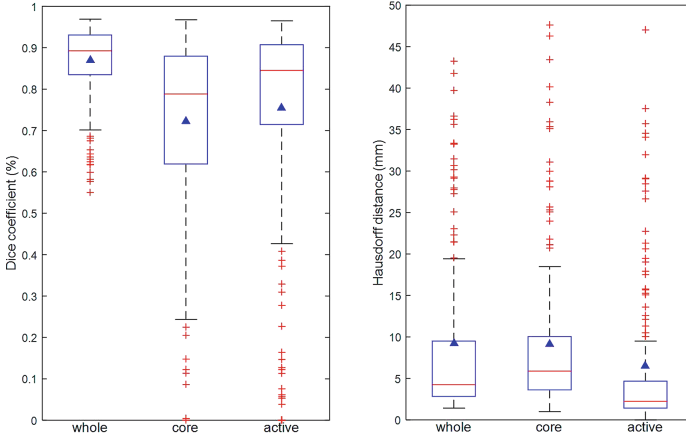


Fig. 8. Boxplots of Dice and Hausdorff distance on 274 cases. The blue triangle shows the mean and red line shows the median. (Color figure online)

Table 2. Comparison of the average dice scores of proposed method and top performing algorithms in BraTS 2016 Challenge. Note that not all the top performing algorithms report their results on all the 274 cases.

	Whole	Core	Active	Data
Chang [4]	87	81	72	29 HGG, No LGG
Kamnitsas [8]	89.6	76.3	72.4	All 220 HGG, 54 LGG
Zeng [15]	89	77	65	176 HGG, 54 LGG
Le Folgoc [6]	82	73	75	176 HGG, no LGG
Meier [10]	84.7	64.7	70	MEDIAN on the BRATS-2015 TESTING database
Our	87.0	72.2	75.4	Whole+Core 220 HGG, 54 LGG. “Active” on HGG

2016 Testing Phase. We apply our model trained with BraTS 2016 training data set on the BraTS 2016 testing data set (291 cases). The results on testing data set are not as good as the ones we get on the training data set. One possible reason is that the cases of the training data set are all of 1.5 T while testing data set contains many 3 T cases.

4 Conclusion

In this paper, we considered the problem of fully automatic multimodal brain tumor segmentation. We proposed a two-stage approach for efficiency in terms of both time and memory. We first utilized anatomy structure information to

normalize data intensity and segment the whole tumor. Next, we employed the random forest classifier on the voxels within the dilated initial segmentation for multi-class tumor segmentation. A novel pathology-guided refinement was applied to further improve accuracy. Promising results are shown on BraTS 2015 training dataset.

References

1. The Virtual Skeleton Database (VSD). www.virtualskeleton.ch
2. Bezdek, J.: A convergence theorem for the fuzzy ISODATA clustering algorithms. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI* **2**(1), 1–8 (1980)
3. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
4. Chang, P.D.: Fully convolutional neural networks with hyperlocal features for brain tumor segmentation. In: *Proceedings MICCAI-BRATS Workshop 2016*, pp. 4–9 (2016)
5. Dunn, J.: A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *J. Cybern.* **3**(3), 32–57 (1973)
6. Folgoc, L.L., Nori, A.V., Alvarez-Valle, J., Lowe, R., Criminisi, A.: Segmentation of brain tumors via cascades of lifted decision forests. In: *Proceedings MICCAI-BRATS Workshop 2016*, pp. 26–30 (2016)
7. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. In: *Proceedings of MICCAI (2011)*
8. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A., Criminisi, A., Rueckert, D., Glocker, B.: DeepMedic on brain tumor segmentation. In: *Proceedings MICCAI-BRATS Workshop 2016*, pp. 18–22 (2016)
9. Kleesiek, J., Biller, A., Urban, G., Köthe, U
10. Meier, R., Knecht, U., Wiest, R., Reyes, M.: CRF-based brain tumor segmentation: alleviating the shrinking bias. In: *Proceedings MICCAI-BRATS Workshop 2016*, pp. 35–39 (2016)
11. Menze, B.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015)
12. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egab, A., Yushkevich, P.A., Gee, J.C.: N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**(6), 1310–1320 (2010)
13. Tustison, N., Wintermark, M., Durst, C., Avants, B.: ANTs and Árboles. In: *Proceedings MICCAI-BRATS Workshop 2013*, pp. 47–50 (2013)
14. Vezhnevets, V., Konouchine, V.: “GrowCut” - interactive multi-label N-D image segmentation by cellular automata. In: *Proceedings of GraphiCon*, pp. 150–156 (2005)
15. Zeng, K., Bakas, S., Sotiras, A., Akbari, H., Rozycki, M., Rathore, S., Pati, S., Davatzikos, C.: Segmentation of gliomas in pre-operative and post-operative multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework. In: *Proceedings MICCAI-BRATS Workshop 2016*, pp. 60–67 (2016)