Proceedings of MICCAI-BRATS 2016

Multimodal Brain Tumor Image Segmentation Benchmark: “Change Detection”

http://www.braintumorsegmentation.org/

Introduction:

Munich, August 2016

Because of their unpredictable appearance and shape, segmenting brain tumors from mult-modal imaging data is one of the most challenging tasks in medical image analysis. There is, however, a growing body of literature on computational algorithms addressing this important clinical task, with several dozen of new methods published in the past three years alone. Unfortunately, open data sets for designing and testing these algorithms are not currently available, and private data sets differ so widely that it is hard to compare the different segmentation strategies that have been reported so far. Critical factors leading to the large variety in training and test data include the imaging modalities employed (T1 and T2 MRI tissue contrasts, T2 FLAIR and T1 contrast-enhanced MRI, various parametric maps from diffusion or perfusion imaging, or any subset of these); the type of the lesion (primary or secondary tumors, solid or infiltratively growing); and the state of disease (images may not only be acquired prior to treatment, but also afterwards then showing different signs of radio-therapy and cavities from resection).

In order to gauge the current state-of-the-art in automated brain tumor segmentation and compare different methods, we organized "Multimodal Brain Tumor Image Segmentation" (BRATS) MICCAI challenges since 2012. As part of this challenge we generated an annotated data set with about 60 low and high grade cases. The data are publicly available from the VSD[1] and the MIDAS[2] web pages, two online platforms for hosting and evaluating image segmentation benchmarks. In 2012 and 2013, altogether 20 groups participated in the challenges, a related journal manuscript[3] describing the results of these efforts has recently been published by TMI. In 2014 we continued BRATS at MICCAI in Boston, also presenting a new data set primarily generated using image data of The Cancer Imaging Archive (TCIA)[4] that we also used during BRATS 2015 in Munich.

At MICCAI 2016 we continue the MICCAI-BRATS challenge, emphasizing longitudinal segmentation tasks and asking the participants to estimate not only the size of relevant tumor structures, but also to predict whether the tumor was "progressing", "shrinking", or remained stable for a set of two scans of a given patient.

This volume reports short descriptions of the algorithms of 16 groups who indicted interest in participating in BRATS 2016. After the conference, participants will be invited to contribute long versions of these articles to the LNCS volume of the MICCAI Brain Lesion Segmentation (BrainLes) workshop.

[1] https://www.smir.ch/BRATS/Start2015
[2] http://challenge.kitware.com/midas/folder/102
[3] http://dx.doi.org/10.1109/TMI.2014.2377694
[4] http://www.cancerimagingarchive.net/

The organizers:

Bjoern Menze
Technische Universität München

Mauricio Reyes
University of Bern

Jayashree Kalpathy-Cramer
MGH Harvard

Keyvan Farahani
National Cancer Institute NIH

Spyridon Bakas &
Christos Davatzikos
University of Pennsylvania

Contributions:

[2nd Edition of the proceedings, updated October 2016]

Semi-Automatic Brain Tumor Segmentation Using Support Vector Machines and Interactive Seed Selection

Mustafa Arikan[1], Bernhard Fröhler[1], Torsten Möller[1]

[1] University of Vienna, Visualization and Data Analysis Research Group

Abstract. Brain tumor segmentation using multi-modal MRI data sets is important for the evaluation of the progression of the disease and the effects of a chosen treatment strategy. In this article we present a semi-automatic segmentation approach using support vector machines trained on a set of interactively selected seeds. Preliminary results on the training data show the importance of senseful selection of seeds for high segmentation accuracy.

1 Introduction

Gliomas are brain tumors in adults, presumably originating from glial cells and infiltrating the surrounding tissues. Gliomas are divided into Low Grade Gliomas (LGG) and High Grade Gliomas (HGG), which are the more aggressive form of the disease. For both variants the evaluation of the progression of the disease and the effects of a chosen treatment strategy are done using imaging modalities from MRI[1]. The goal of our approach is the segmentation of brain tumor regions in multi-modal MRI data sets. The tumor regions are defined through intensity changes relative to the surrounding normal tissue [1]. Intensities in the MRI modalities do not have a tissue specific value, meaning that the same tissue has a wide range of intensities [2]. The MRI data sets come from different scanners with different configurations, which make the segmentation more challenging.

2 Method

In a first step, we use an anisotropic diffusion filter for noise reduction. In the next step a bounding box containing the tumor regions is extracted in all modality data sets. In the third step we randomly select a certain number of seeds in the data set. The class labels of these seeds are interactively determined within the user interface. For the purpose of segmentation using support vector machines [3] we use Flair, T1, T1c and T2 modalities and the spatial position of the seeds. The intensity values are scaled between -1 and 1. To find good parameters for the SVM we apply a Latin hypercube sampling strategy [4].

3 Experiments and Results

Our method is validated on four selected examples from the BRATS 2015 training data sets [5].

Table 1. Dice values for selected examples

Training data set	Dice
brats_tcia_pat222_0122	0.813332
brats_tcia_pat226_0090	0.775794
brats_tcia_pat260_0001	0.753066
brats_2013_pat0015_1	0.893903

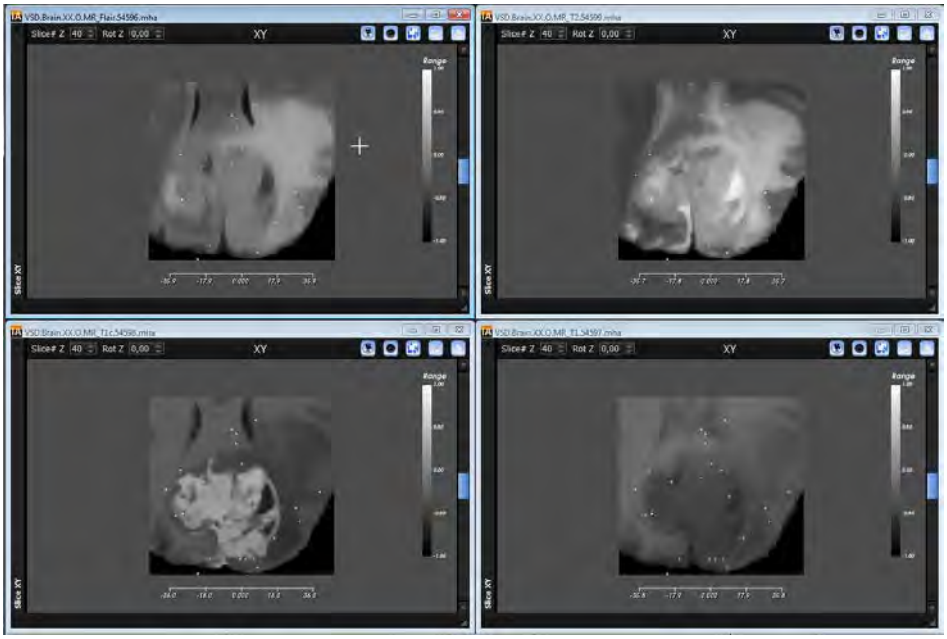


Fig.1: The four images show the MRI modalities used as input data sets to the SVM model and the selected seeds: green for label 0, yellow for label 1, violet for label 2, red for label 3 and blue for label 4.

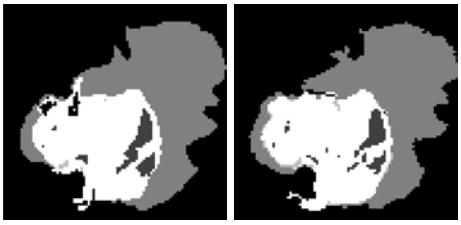


Fig.2: The two images show the segmented result (left) and the ground truth (right) for the example brats_2013_pat0015_1.

4 Conclusion

In this paper, we proposed a brain tumor segmentation method based on support vector machines and interactive seed selection. Our method shows the importance of well selected seed points to achieve high segmentation accuracy. In the future, we intend to extend the method to guide the expert to select the important seed points with constant feedback.

References

1. B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," in IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993-2024, Oct. 2015.
2. Jansen, M. J. A., Evaluation of intensity normalization methods for MR images, MSc report, University of Twente, 2015
3. C.-C. Chang and C.-J. Lin. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.
4. B. Fröhler, T. Möller, C. Heinzl: GEMSe: Visualization-Guided Exploration of Multi-channel Segmentation Algorithms. Computer Graphics Forum, Vol. 35, No. 3, June 2016, 191-200.
5. B. Menze, M. Reyes, and K. Van Leemput. The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging, (99), 2014.

Fully Convolutional Neural Networks with Hyperlocal Features for Brain Tumor Segmentation

Peter D. Chang, MD[1]

[1]Columbia University Medical Center, Department of Radiology, New York City, New York, USA

Abstract. In this paper, a novel yet simple convolutional neural network architecture is introduced uniquely adapted to the challenges of medical image segmentation, employed here in the setting of brain tumors. The network is implemented with a fully convolutional architecture, with a dense output classification matrix equal in size to the input image (240 x 240 voxel predictions for each forward pass). Importantly the network combines activations from the deepest convolutional output layer with hyperlocal features from the original input image just prior to final segmentation. The network is composed of only 130,400 parameters, and given its fully convolutional architecture can complete segmentation for an entire brain volume in less than one second. Despite this simple network design, the model exceeds state-of-the art with complete tumor, core tumor and enhancing tumor validation Dice scores of 0.87, 0.81 and 0.72 respectively.

adfa, p. 1, 2011.

1 Introduction

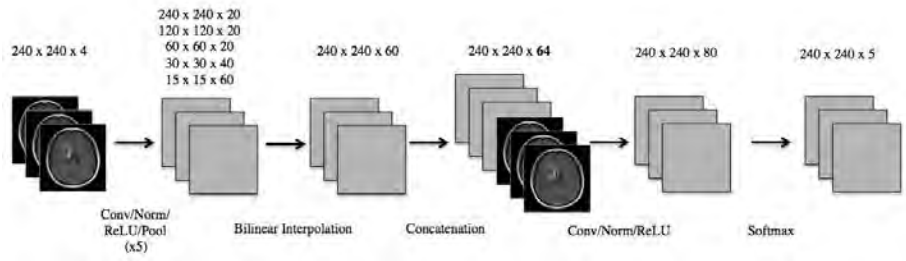
In recent years, convolutional neural networks (CNN) have become the technique of choice for state-of-the-art implementations in various image classification task (1–5). The top finisher of the BraTS 2015 challenge was the first to apply CNN to brain tumor segmentation (6). In this paper, the conventional CNN architecture commonly used for generic segmentation tasks is adapted for medical imaging in several ways.

First, instead of patch-wise classification of the center pixel, the method proposed here uses a fully convolutional architecture with a dense output classification matrix equal in size to the input image (an entire axial MR slice of 240 x 240 voxel predictions for each forward pass). This method was first introduced by Long et al (7), and requires upsampling of the deepest layers to the original input size. The advantages of this architecture are two-fold: (1) image features of the deepest layer represent contextual anatomical cue beyond that which may be captured by a small "patch" of the input image; and (2) the speed of implementation whereby a single forward pass results in classification of all voxels within the input axial slice.

From the perspective of a board-certified radiologist, an additional key insight can be used to further improve this fully convolutional architecture: once contextual anatomic cues have been extracted from the image, final segmentation is heavily influenced by the signal intensity of that particular voxel (as opposed to textural or other high-order features). That is, once the general location of a tumor has been identified, the decision of where to demarcate tissue boundaries is almost exclusively based on voxel brightness. This hyperlocal quality of medical image segmentation can be captured with a direct-acyclic graph (DAG-CNN) architecture that re-introduces the original input image into the network just two layers prior to final classification (8).

2 Convolutional Neural Network Architecture

Fig. 1. CNN architecture (130,400 parameters total)



2.1 Conventional Convolutional Networks

Convolutional neural networks are an adaption of the traditional artificial neural network architecture whereby stacks of convolutional filters kernels and nonlinear activation functions act as a mapping function to transform a multidimensional input image into a desired output (9). The CNN used in this study uses 5 stacks of 5 x 5 filters followed by the ReLU nonlinear function (Fig. 1). Batch normalization is used between the convolutional and ReLU layers to limit drift of layer activations during training (10). The ReLU activation layers are followed by a max-pooling operation with 2 x 2 window size and a stride length of 2 voxels, effectively reducing each channel of activations by 75 percent (11). Similarly, the number of activation channels in each subsequent layer is progressively increased from 20 to 40 to 60, reflecting increased feature complexity.

2.2 Fully Convolutional Networks

To implement a fully convolutional architecture, the deepest layer of feature activations (after five serial convolutional and nonlinear activation operations) is upsampled to the original input image size by simple bilinear interpolation. A convolutional transpose operation with learnable weights was considered as an alternative however given the large number of filter parameters in such an operation, this option resulted in significant over-fitting of the data.

2.3 Hyperlocal Feature Concatenation

The original input data is reintroduced into the network by means of a concatenation operation across channels. The resulting concatenated layer is composed of sixty-four channels, four corresponding to the original image (pre-contrast, post-contrast, FLAIR and T2) and sixty additional inputs from the deepest layer representing various high order features. To limit influence from surrounding voxels, the final classification is based on two narrow 3 x 3 convolutional filters, which act to capture the optimal combination of general and hyperlocal features to produce the segmentation map.

2.4 Training

Training is implemented using standard stochastic gradient descent technique with Nesterov momentum (12). Parameters are initialized using the heuristic described by He et al (13). L2 regularization is implemented to prevent over-fitting of data by limiting the squared magnitude of the kernel weights. To account for training dynamics, the learning rate is annealed and the mini-batch size is increased whenever training loss plateaus. Furthermore a normalized gradient algorithm is employed to allow for locally adaptive learning rates that adjust according to changes in the input signal (14).

2.5 Image Preprocessing

Each individual input image slice was processed with an intensity normalization algorithm by means of histogram matching (15). The reference template histogram used in histogram matching was generated by pooling all the image data used in this study, and subsequently scaled from [0,1]. No other preprocessing methods were utilized.

2.6 Implementation Details

Software code for this study was written in Matlab (R2015b), and benefited greatly from the MatConvNet toolbox (16). Experiments were performed on a GPU-optimized workstation with a single NVIDIA GeForce GTX Titan X (12GB). The combined software and hardware configured allowed mini-batches of 64 input images be processed in approximately 0.52 seconds. An entire brain volume with 151 axial slices could be classified in approximately 0.93 seconds.

3 Experiments and Results

All 144 patients with high-grades gliomas (HGG) in the BRATS 2016 database were included in this study. Eighty percent of the patients (n = 115) were randomly assigned to the training data set, with the remaining 20 percent of the patients (n = 29) used as an independent validation set. All four channels across an entire axial cross-sectional image slice (240 x 240 x 4 voxels) were used as inputs into the CNN. Given this setup, a total of 17,365 images were used for training and 4,379 images were used for validation.

Training proceeded for a total of 50 epochs before reaching convergence. Dice scores and Hausdorff distances (mean and range) for the validation set data are reported below.

Table 1. Validation Set Accuracy

	Complete Tumor	*Core Tumor*	*Enhancing Tumor*
Dice	0.87 (0.75-0.93)	0.81 (0.63-0.88)	0.72 (0.37-0.88)
Hausdorff (mm)	9.1 (3.0-27)	10.1 (4.2-28)	6.0 (3.2-22)

4 Discussion and Conclusion

The CNN architecture presented in this paper is simple, efficient and highly optimized to the task of medical image segmentation. Despite the small number of learned parameters, the final network exceeds state-of-the-art in the challenging brain tumor segmentation task, and is able to complete segmentation for an entire volume in less than one second on a GPU optimized workstation.

References

1. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. 2012. p. 1097–105.
2. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014.
3. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions. 2014 Sep 16 [cited 2016 Jul 31]; Available from: http://arxiv.org/abs/1409.4842
4. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015 Dec 10 [cited 2016 Jul 31]; Available from: http://arxiv.org/abs/1512.03385
5. Ciresan D, Giusti A. Deep neural networks segment neuronal membranes in electron microscopy images. Adv neural … [Internet]. 2012;1–9. Available from: http://papers.nips.cc/paper/4741-deep-neural-networks-segment-neuronal-membranes-in-electron-microscopy-images
6. Havaei M, Davy A, Warde-Farley D. Brain Tumor Segmentation with Deep Neural Networks. arXiv Prepr arXiv … [Internet]. 2015;13. Available from: http://arxiv.org/abs/1505.03540
7. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. 2015 IEEE Conf Comput Vis Pattern Recognit [Internet]. 2014;3431–40. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7298965\nhttp://arxiv.org/abs/1411.4038
8. Yang S, Ramanan D. Multi-Scale Recognition With DAG-CNNs. 2015. p. 1215–23.
9. Lecun Y, Bengio Y, 4g332 R. Convolutional Networks for Images, Speech, and Time-Series LeCun & Bengio: Convolutional Networks for Images, Speech, and Time-Series LeCun & Bengio: Convolutional Networks for Images, Speech, and Time-Series.
10. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv:150203167 [Internet]. 2015 [cited 2016 Jul 31];1–11. Available from: http://arxiv.org/abs/1502.03167
11. Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines.
12. Bengio Y, Boulanger-Lewandowski N, Pascanu R. Advances in optimizing recurrent networks. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. 2013. p. 8624–8.
13. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 Feb 6 [cited 2016 Jul 31]; Available from: http://arxiv.org/abs/1502.01852
14. Mandic DP. A Generalized Normalized Gradient Descent Algorithm. IEEE Signal Process Lett [Internet]. IEEE; 2004 Feb [cited 2016 Jul 31];11(2):115–

8. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1261952
15. Shapira D, Avidan S, Hel-Or Y. Multiple Histogram Matching. Image Processing (ICIP), 2013 20th IEEE International Conference on. 2013. p. 2269–73.
16. Vedaldi A, Lenc K. MatConvNet. Proc 23rd ACM Int Conf Multimed - MM '15 [Internet]. New York, New York, USA: ACM Press; 2015 [cited 2016 Jul 31];689–92. Available from: http://dl.acm.org/citation.cfm?doid=2733373.2807412

Assessing the Non-Negative Matrix Factorization Level Set Segmentation on the BRATS Benchmark

Dimah Dera[1], Nidhal Bouaynaya[1], and Hassan M Fathallah-Shaykh[2]

[1] Department of Electrical and Computer Engineering, Rowan University, Glassboro, NJ, USA
[2] Department of Neurology and Mathematics, University of Alabama at Birmingham, Birmingham, AL, USA

Abstract. We present a novel fully automated pixel-wise accurate and robust method for image segmentation, termed NMF-LSM [5]. Non-negative Matrix Factorization (NMF) is a decomposition technique that reduces the dimensionality of the image by extracting its distinct regions. When combined with the level-set method (LSM), NMF-LSM proves an efficient method for image segmentation. We apply NMF-LSM segmentation on the training data of the Multimodel Brain Tumor Image Segmentation Benchmark (BRATS) [7]. By evaluating the performance of the segmentation results with the ground truth using Jaccard similarity (JS), Dice score and root mean square error (RMSE), we show that NMF-LSM is highly accurate in segmenting the complex tumor structures (edema, non-enhancing solid core, necrotic/cystic core and enhancing core) in multi-contrast MR scans of low- and high-grade glioma patients. NMF-LSM yields high JS (mean = 0.86) and Dice (mean = 0.91) scores and low values of RMSE (mean = 0.08). The method is computationally efficient as the computational time for the segmentation of 465 images is 240 minutes on a personal PC and 60 minutes on a supercomputer.

1 Introduction

Non-negative matrix factorization (NMF) has shown promise as a robust clustering and data reduction technique in DNA microarrays clustering and classification [2] and learning facial features [6]. NMF is distinguished from the other methods, such as principal components analysis and vector quantization, by its use of non-negativity constraints. These constraints lead to a parts-based representation because they allow only additive, not subtractive, combinations. The first use of NMF for image segmentation was pioneered by our group in [4].

The level set method (LSM) is one of the most powerful and advanced methods to extract object boundaries in computer vision [3], [8]. The basic idea of LSM is to evolve a curve in the image domain around the object or the region of interest until it locks onto the boundaries of the object. The level set approach represents the contour as the zero level of a higher dimensional function, referred to as the "level set function". The segmentation is achieved by minimizing a functional that tends to attract the contour towards the objects features. The advantages of NMF-LSM is that it: (i) is robust to noise, initial condition, and intensity inhomogeneity because it relies on the distribution of the pixels rather than the intensity values and does not introduce spurious or nuisance model parameters, which have to be simultaneously estimated with the level sets, (ii) uses NMF to discover and identify the homogeneous image regions, (iii) introduces a novel spatial functional term that describes the local distribution of the regions within the image, and (iv) is scalable, i.e., able to detect a distinct region as small as desired.

In this paper, we apply our novel NMF-LSM approach [5], on the training data of the Multimodel Brian Tumor Image Segmentation Benchmark (BRATS) [7]. By comparing the segmentation results with the ground truth using Jaccard similarity (JS), Dice score and root mean square error (RMSE), we show that NMF-LSM achieves notable high values of JS and Dice score and low values of RMSE.

2 NMF-based Clustering

The NMF-LSM method is fully automated, pixel-wise accurate, less sensitive to the initial selection of the contour(s) or initial conditions compared to state-of-the-art LSM approaches, robust to noise and model parameters, and able to detect as small distinct regions as desired (for a detailed description, please see [5]). These advantages stem from the fact that NMF-LSM method relies on histogram information instead of intensity values and does not introduce nuisance model parameters. We have applied the NMF-LSM method to analyze the MRIs of two patients with grade 2 and 3 non-enhancing oligodendroglioma and compared its measurements to the diagnoses rendered by board-certified neuroradiologists; the results demonstrate that NMF-LSM can detect earlier progression times and may be used to monitor and measure response to treatment [5].

3 Simulation Results and Discussion

We evaluate the performance of the NMF-LSM on the training data of the BRATS Benchmark [7]. According to [7], we are interested in segmenting the following tumor structures: (i) the "whole" tumor region visible in FLAIR (including all tumor structures, edema, non-enhancing (solid) core, enhancing core and necrotic or fluid-filled core), (ii) the tumor "core" region visible in T2 (including all tumor structures except "edema"), and (iii) the "active" tumor region visible in T1c (only containing the "enhancing core" structures that are unique to high-grade cases). Figure 1 shows examples from the BRATS training data, with the ground truth of the tumor regions as inferred from BRATS benchmark (blue lines) and NMF-LSM segmentation result (magenta lines). Each row shows two cases of high-grade tumor (rows 1-2) or low-grade tumor (rows 3-4), showing for each case: FLAIR with outlines of the whole tumor region (left); T2 with outlines of the core region (center); T1c with outlines of the active tumor region (right).

The quantitative evaluation of the segmentation accuracy will be assessed using three different similarity measures, Jaccard Similarity (JS) [9], Dice Score [1], and root mean square error (RMSE). Although Jaccard and Dice scores are very similar, the Jaccard similarity is more sensitive when the regions are more similar, while the Dice score gives more weighting to instances where the two images agree. Both of the JS and Dice scores provide values ranging between 0 (no overlap) and 1 (perfect agreement). The root mean square error RMSE is a distance measure that gives the difference between two image regions or image intensities.

We assess eight cases of multi-contrast MR scans, 4 for low-grade and 4 for high-grade glioma patients from the BRATS benchmark. In each case, we compare the segmentation results with the ground truth for the three different tumor regions ("whole" tumor, tumor "core" and "active" tumor) using the three different MRI scans available in the training data (FLAIR, T2 and T1c).

By looking at Fig. 2, we see that NMF-LSM achieves high values of the Jaccard similarity and Dice score, and low values of RMSE for the three different regions. For the "whole" tumor region, the JS range is (0.76 - 0.97) with mean 0.87 and standard deviation 0.06. The dice score range is (0.84 - 0.96) with mean 0.91 and standard deviation 0.04. The RMSE range is (0.06 - 0.13) with mean 0.09 and standard deviation 0.02, (see Fig. 2a).

For the tumor "core" region, the JS range is (0.75 - 0.97) with mean 0.87 and standard deviation 0.05. The dice score range is (0.82 - 0.96) with mean 0.91 and standard deviation 0.03. The RMSE range is (0.05 - 0.15) with mean 0.08 and standard deviation 0.03, (see Fig. 2b). For the "active" tumor region, the JS range is (0.73 - 0.97) with mean 0.84 and standard deviation 0.06. The dice score range is (0.79 - 0.95) with mean 0.87 and standard deviation 0.04. The RMSE range is (0.04 - 0.09) with mean 0.07 and deviation deviation 0.01, (see Fig. 2c).

In summary, NMF-LSM achieves high values (close to 1) of JS and Dice scores when compared against the ground truth, while it achieves low values (close to 0) of RMSE, for the three specified tumor structures ("whole" tumor, tumor "core" and "active" tumor). The NMF-LSM model converges remarkably faster than other state-of-the-art segmentation approaches [5]. For the BRATS data, the average

computation time per case (FLAIR, T1c, and T2 MRI scan around 465 images) is around 240 min on a personal PC and 60 min on a supercomputer which is computationally efficient. This property can be explained by the fact that the NMF-LSM estimates less parameters than other state-of-the-art approaches [5].

4 Conclusion

In this paper, we applied our novel NMF-LSM segmentation approach in [5] on the training data of the Multimodel Brain Tumor Image Segmentation Benchmark (BRATS) [7]. The quantitative evaluation of the segmentation accuracy using JS, Dice score, and RMSE have shown the accuracy of the NMF-LSM in segmenting complex tumor structures (edema, non-enhancing solid core, necrotic/cystic core and enhancing core) in multi-contrast MR scans of low- and high-grade glioma patients. NMF-LSM achieves notable high values of JS with mean 0.86 and Dice score with mean 0.91 and low values of RMSE with mean 0.08 over the three required regions ("whole" tumor, tumor "core" and "active" tumor).

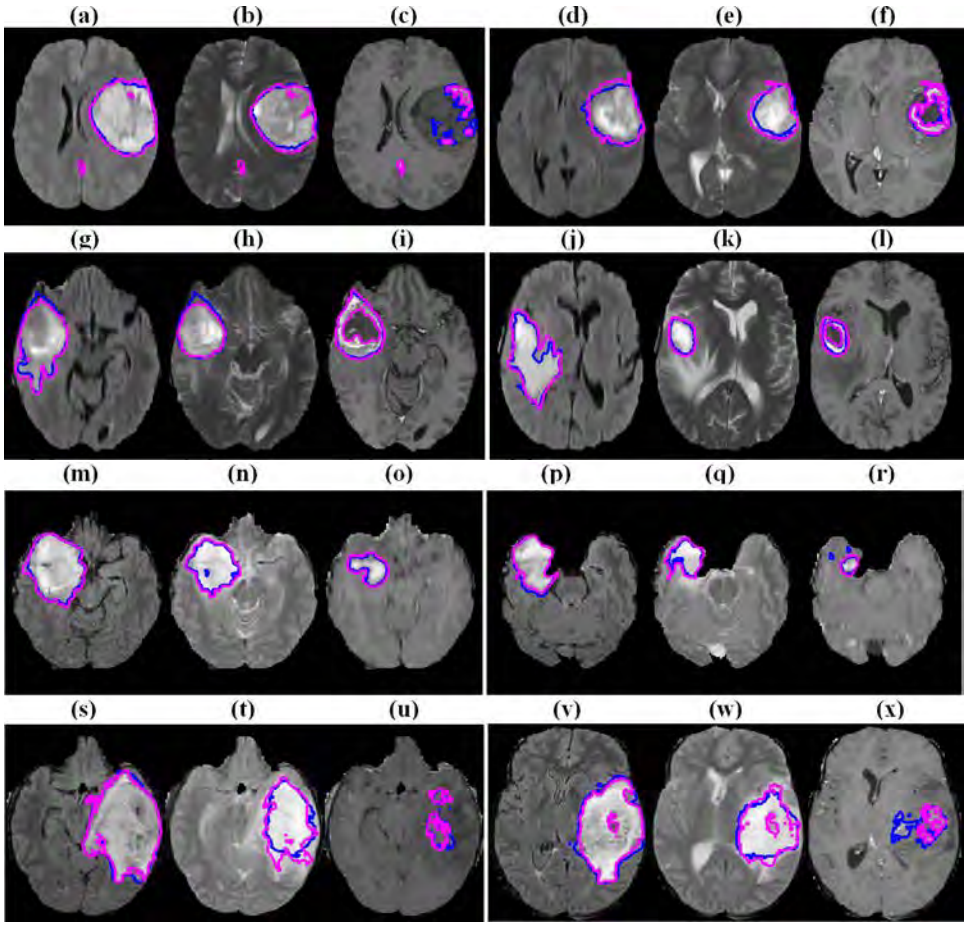


Fig. 1. Examples from the BRATS training data, with the ground truth of the tumor regions as inferred from BRATS benchmark (blue lines) and NMF-LSM segmentation result (magenta lines).

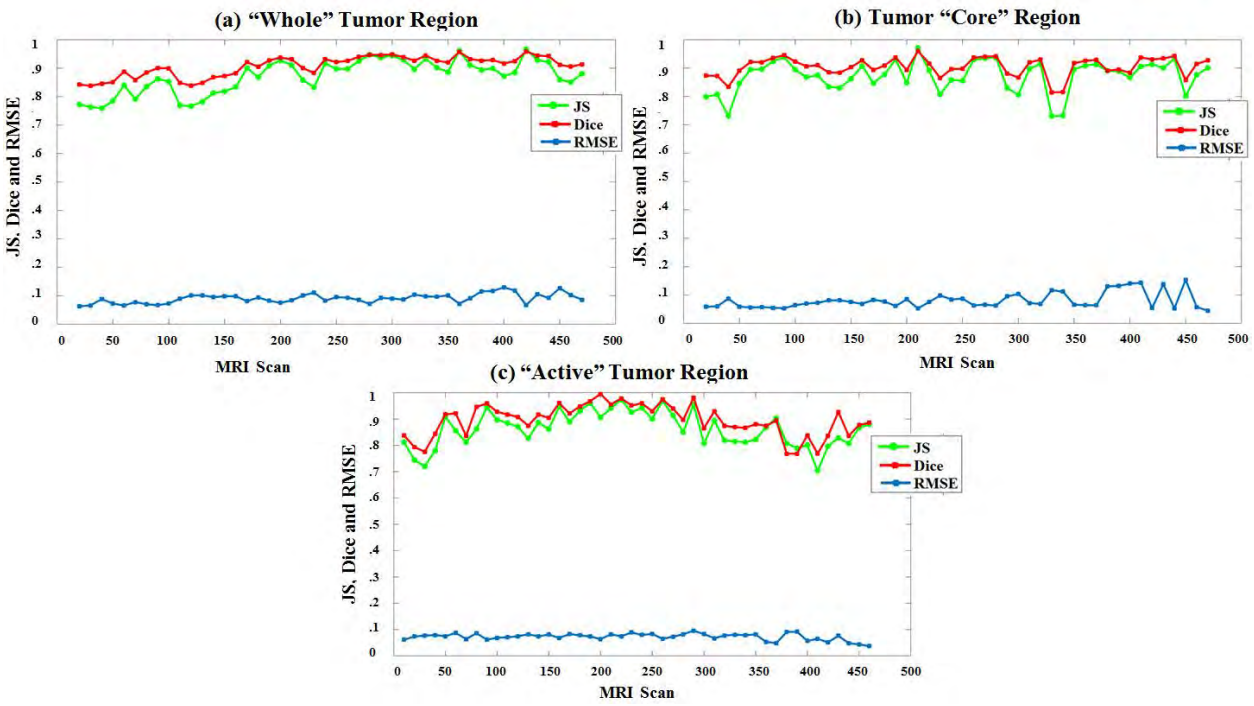


Fig. 2. Performance evaluation of the proposed NMF-LSM approach on the training data provided by BRATS benchmark using JS, Dice score and RMSE.

References

1. Babalola, K.O., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., Smith, S., Cootes, T.F., Jenkinson, M., Rueckert, D.: Comparison and evaluation of segmentation techniques for subcortical structures in brain mri. Medical Image Computing and Computer-Assisted Intervention 5241, 409–416 (2008)
2. Bayar, B., Bouaynaya, N., Shterenberg, R.: Probabilistic non-negative matrix factorization: Theory and application to microarray data analysis. Journal of Bioinformatics and Computational Biology 12, 25 (2014)
3. Cohen, L.D., Cohen, I.: Finite-element methods for active contour models and balloons for 2-d and 3-d images. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), 1131–1147 (1993)
4. Dera, D., Bouaynaya, N., Fathallah-Shaykh, H.M.: Level set segmentation using non-negative matrix factorization of brain mri images. IEEE International Conferenceon Bioinformatics and Biomedicine (BIBM), Washington, D.C. pp. 382 – 387 (2015)
5. Dera, D., Bouaynaya, N., Fathallah-Shaykh, H.M.: Automated robust image segmentation: Level set method using non-negative matrix factorization with application to brain mri. Bulletin of Mathematical Biology 78, 1 – 27 (2016)
6. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–793 (1999)
7. Menze, B., et al: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging 34 (2015)
8. Tsai, A., Yezzi, A., S.Willsky, A.: Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation and magnification. IEEE Transactions on Image Processing 10, 1169–1186 (2001)
9. Vovk, U., Pernus, F., Likar, B.: A review of methods for correction of intensity inhomogeneity in mri. IEEE Transactions on Medical Imaging 26, 405–21 (2007)

Brain Tumor Segmantation using Random Forest trained on iterative selected patients

Abdelrahman Ellwaa, Ahmed Hussein, Essam AlNaggar, Mahmoud Zidan, Michael Zaki, Mohamed A. Ismail, and Nagia M. Ghanem

Faculty of Engineering, Alexandria University

Abstract. In this paper we present a fully automated method for brain tumor segmentation in multi-modal magnetic resonance images(MRI) which is very important for surgery and treatment planning and not time consuming as the manual methods. We introduce an iterative random forest approach which tries to improve its accuracy by iteratively choosing the best patients to use it in training the random forest. The experiments on the data set (2016 BRATS) show that the methods performance is comparable to the current state-of-the-art.

Keywords: Brain Tumor Segmentation, Random Forests

1 Introduction

Brain tumor segmentation is a challenging problem as it imposes an unpredictable behavior and it doesn't appear in a fixed location or has a specific size or shape. The tumor intensity differs from one patient to another. In the HGG patients the tumor consists of enhancing, non-enhancing and necrotic parts, while in the LGG patients it is not necessarily to include an enhancing part.

In this paper we introduce a random forest approach which chooses the patients to be used in training according to cost function, training is iterative at each iteration some patients are added to the training set to be used in the next step, this approach tries to prevent over fitted random forest by choosing patients that gets the worst results in the previous iteration. Through the paper we will illustrate in details the approach and the parameters of the approach.

2 Method

The method used in different models consists of four major stages, pre-processing, feature extraction and selection, training random forest, and finally a post processing step. Before explaining out the training pipeline, BRATS 2016 dataset was partitioned into training dataset (70%), testing dataset (20%), and validation dataset (10%).

II

2.1 Preprocessing

The first step in pre-processing is applying a bias field correction on the MR images which is used to correcting the intensity non-uniformity in MR images and this is applied using open source code N4ITK [1]. The second step in pre-processing is histogram matching Normalization [2], [3].

2.2 Feature extraction and selection

In this phase we extract 328 features from the MR images after being pre-processed. Mainly, three categories of features have been extracted, gradient features, appearance features, and context aware features. Most of them were from other published papers form BRATS challenge [4], [5], [6]. The first type of features is gradient features which includes gradient filter at Sigma 0.5, 1, 2, 3 for x, y, z, and their resultant, difference gradient features, Laplace features, and Recursive Gaussian features. The second type of features is Appearance features which includes the voxel intensity, its logarithmic and exponential transformation. The last type of features is the context aware features, which is intensity based, it includes features extracted from the neighbouring voxels, the surrounding cube of the voxel, as most similar, most different, Minimum, Maximum, Range, Kurtosis, Skewness, Standard deviation, and Entropy for all modalities, as well as the Local Histogram is building the histogram of the cube surrounding the voxel and partitioned into eleven bins, all the previous features were extracted for all modalities Flair, T1, T1c, and T2. Then feature selection techniques were applied to reduce the features and select the most discriminative features to be used in training, testing and validating the random forest parameters. Since our main classification method is Random forest so we exploit it in our feature selection phase, such that one of the main steps of random forest algorithm is to compute the features' importance.

2.3 Training Random Decision Forest

H2O [7] tool was used in training the random forest, random forest parameters were selected according to validation step, the best configuration found to be using a random forest with number of trees equals to 45 and number of nodes to split on equals to the square root of the number of features.

2.4 Post processing

The post-processing step is applying binary morphology filter to the output image of the classifier.The main target of applying this filter is connecting large tumorous regions and remove small isolated regions. The radii used in binary morphology were validated using the validation dataset.

3 Experiment and Results

3.1 Experiment

This section explains the models used in classification.

Iterative Model The iterative model mainly addresses the problem of choosing a subset of training patients to train the random forest, so the model is trained through number of iterations, in each iteration the number of patients increases according to a cost function so the worst n patients are added, there is also maximum number of patients that is selected according to the hardware resources available. The flow chart in figure 1 explains how the iterative model works.

There are many parameters that must be specified first as the number of patients added in each iteration that is set to 5 patients per iteration, initial set of patients that are 30 patients (2013 dataset 20 HGG and 10 LGG), maximum number of LGG patients to prevent overfitting LGG patients that is set to 18 patients, the maximum number of patients that is limited to 50 patients, and the cost function which is equals to

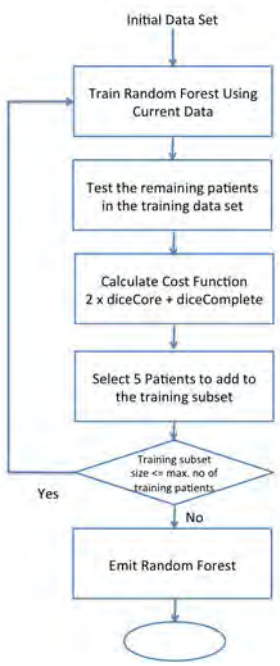


Fig. 1. Iterative Model Flowchart

$$costFuction = 2 * coreDice + completeDice$$

One-phase Model and Cascade Model Both models are trained on 50 randomly chosen patients from the training data set, the random forest consists of 100 trees each of depth 45, The first model classifies the five labels using one random forest, the second model uses two random forests one to classify tumor and non-tumor and the second random forest to classify tumorous cells to 4 labels.

3.2 Results

Those models are tested on 20 (15 HGG and 5 LGG) unseen patients, the results are shown in table 1 by the dice, specificity and sensitivity scores.

IV

	complete			core			enhancing		
Model	Dice	Sens	Spec	Dice	Sens	Spec	Dice	Sens	Spec
One-phase HGG	0.81 ± 0.09	0.84 ± 0.10	0.98 ± 0.01	0.62 ± 0.23	0.83 ± 0.16	0.98 ± 0.01	0.75 ± 0.15	0.81 ± 0.12	1.00 ± 0.00
One-phase LGG	0.80 ± 0.06	0.76 ± 0.14	0.99 ± 0.01	0.53 ± 0.17	0.74 ± 0.20	0.99 ± 0.01	0.36 ± 0.32	0.62 ± 0.43	1.00 ± 0.00
Cascaded HGG	0.78 ± 0.10	0.89 ± 0.08	0.97 ± 0.02	0.63 ± 0.23	0.88 ± 0.11	0.98 ± 0.02	0.71 ± 0.18	0.83 ± 0.10	1.00 ± 0.00
Cascaded LGG	0.76 ± 0.11	0.91 ± 0.06	0.98 ± 0.01	0.55 ± 0.20	0.79 ± 0.22	0.99 ± 0.01	0.24 ± 0.31	0.57 ± 0.48	1.00 ± 0.00
Iterative HGG	0.80 ± 0.10	0.84 ± 0.11	0.98 ± 0.01	0.72 ± 0.19	0.79 ± 0.18	0.99 ± 0.01	0.73 ± 0.15	0.75 ± 0.12	1.00 ± 0.00
Iterative LGG	0.84 ± 0.02	0.83 ± 0.09	0.99 ± 0.01	0.72 ± 0.12	0.82 ± 0.14	0.99 ± 0.01	0.39 ± 0.33	0.51 ± 0.42	1.00 ± 0.00

Table 1. Models Results

4 Conclusion

In this paper we proposed an approach based on Random Forest that differs from past years' submissions as we mainly tried to extract as much information as we can from our large dataset (BRATS 2016 dataset). We achieved this by applying our iterative selection method to choose the best patients to train our Random Forest with them and then by extracting as much information as we can then applying feature selection, and our proposed method improves the performance over the cascaded method and over training the RF using randomly selected patients.

References

1. N. Tustison and J. Gee: N4itk: Nicks n3 itk implementation for mri bias field correction. Insight Journal, 2009
2. L. G. Nyu and J. K. Udupa: On standardizing the mr image intensity scale image, vol. 1081, 1999
3. L. G. Nyu and J. K. Udupa: New variants of a method of mri scale standardization. Medical Imaging, IEEE Transactions on, vol. 19, no. 2, pp. 143150, 2000.
4. Hugo Dinis Adriano Pinto, Sergio Pereira and Carlos A. Silva: Random decision forests for automatic brain tumor segmentation in multi-modal mri images.
5. Jean-Marc Peyrat Julien Abinahed Eric Malmi, Shameem Parambath and Sanjay Chawla: CaBS, A Cascaded Brain Tumor Segmentation Approach.
6. Matthias Wilms Oskar Maier and Heinz Handels: Highly discriminative features for glioma segmentation in MR volumes with random forests.
7. http://www.h2o.ai/

DeepMedic on Brain Tumor Segmentation

Konstantinos Kamnitsas[1,2⋆], Enzo Ferrante[1], Sarah Parisot[1], Cristian Ledig[1], Aditya Nori[2], Antonio Criminisi[2], Daniel Rueckert[1], and Ben Glocker[1]

[1] Biomedical Image Analysis Group, Imperial College London, UK
[2] Microsoft Research, Cambridge, UK

Abstract. Accurate automatic algorithms for the segmentation of brain tumours have the potential of improving disease diagnosis, treatment planning, as well as enabling large-scale studies of the pathology. In this work we employ DeepMedic [1], a 3D CNN architecture previously presented for lesion segmentation, which we further extend with residual connections to investigate their effect. We also present a series of experiments on the BRATS 2015 training database for evaluating the robustness of the network when less training data are available or less filters are used, aiming to shed some light on the practical questions about requirements for employing such a system.

1 Introduction

Accurate estimation of the relative volume of tumour's sub-components is critical for monitoring progression, radiotherapy planning, outcome assessment and follow-up studies. For this, accurate delineation of the tumour is required. Manual segmentation poses significant challenges for human experts, both because of the variability of tumour appearance but also because of the need to consult multiple images from different MRI sequences in order to classify tissue type correctly. This laborious effort is not only time consuming but prone to human error and results in significant intra- and inter-rater variability [2].

Automatic segmentation systems aim to provide a cheap and scalable solution. Over the years, automatic methods for brain tumour segmentation have attracted significant attention. Representative early work is the atlas-based outlier detection method of [3]. Segmentation was later solved jointly with the registration to a pathological brain of a healthy atlas by [4], who made use of a tumour growth model and the Expectation Maximization algorithm, and [5] who casted the problem to a joint optimization of two Markov Random Fields (MRFs). The state of the art was raised further by supervised learning methods, initially represented mainly by Random Forests, coupled with models such as Gaussian Mixtures for the extraction of tissue type priors [6], MRFs for spatial regularisation and a variety of engineered features [7].

The recent years saw the success of deep learning, with [8] and [9] being the top performing automatic approaches in BRATS 2014 and 2015 [10], using 3D and 2D Convolutional Neural Networks (CNNs) respectively. The latter approached the accuracy of the winning semi-automatic method of [11]. The fact that the employed models are rather simple in design reveals the high potential of CNNs. The method of [12] also exhibited interesting performance, based on a 3-layered 2D network that separately processes each axial slice. The authors empirically showed that the class bias introduced to a network when training with patches extracted equiprobably from the task's classes can be partially alleviated with

⋆ The first author was partly funded by an Imperial College Phd Scholarship.
Email correspondence to: konstantinos.kamnitsas12@imperial.ac.uk

a second training stage using uniformly extracted patches. [13] used an ensemble of 2D networks to process three orthogonal slices of brain MRI. Finally, [1] showed how carefully engineered, multi-scale 3D CNNs of larger size accomplish high performance while remaining computationally efficient. The work also analysed how the size of the input segments relates to the captured class distribution by the training samples. It was shown that this meta-parameter can be exploited for capturing a distribution of training samples that in practice leads to good performance in a variety of segmentation tasks. The system accomplished state of the art results on stroke lesion segmentation, winning the first position in the SISS-ISLES 2015 challenge [14], brain tumours, traumatic brain injuries and placenta segmentation [15].

2 Method

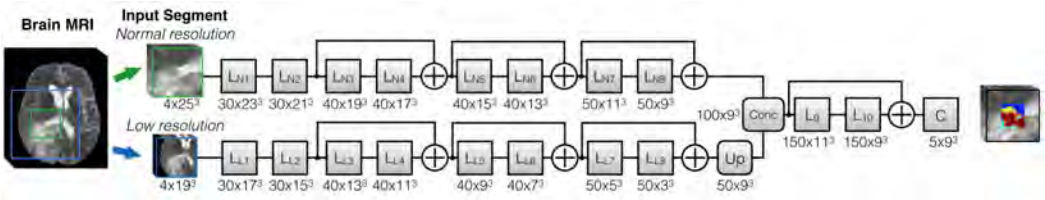


Fig. 1: The DeepMedic [1] extended with residual connections. The operations within each layer block are applied in the order: Batch-Normalization [16], non-linearity and convolution. [17] empirically showed this format leads to better performance. *Up* and *C* represent an upsampling and classification layer respectively. Number of filters and their size depicted as (*Number* × *Size*). Other hyper-parameters as in [1].

DeepMedic is the 11-layers deep, multi-scale 3D CNN presented in [1] for brain lesion segmentation. The architecture consists of two parallel convolutional pathways that process the input at multiple scales to achieve a large receptive field for the final classification while keeping the computational cost low. Inspired by VGG [18], the use of smallest convolutional kernels is adopted, a design choice that the former work showed to be very effective in building deeper and higher performing 3D CNNs while severely limiting the amount of required computations and number of trainable parameters. The CNN is employed in fully a convolutional fashion on image segments in both training and testing stage[4].

We extend the DeepMedic with residual connections in order to examine their effect on segmentation. Residual connections were recently shown to facilitate preservation of the flowing signal and as such have enabled training of very deep neural networks [19,17]. In [19] the authors did not observe a performance improvement when a 18-layers deep network was employed, but only in experiments with architectures deeper than 34-layers. The networks employed on biomedical applications tend to consist of less layers than modern architectures in computer vision. However, the problem of preserving the forward and backwards propagated signal as well as the difficulty of optimization can be substantial in 3D CNNs due to the large number of trainable parameters in 3D kernels, as previously discussed in [1]. For this reason we set off to investigate such an architecture.

We extended the network by adding residual connections between the outputs of every two layers, except for the first two of each pathway to enforce abstracting away from raw intensity values. The architecture is depicted in Fig. 1. In our work, a residual connection after layer l performs the element-wise addition $\oplus$ between corresponding channels of the

[4] Code publicly available at: `https://github.com/Kamnitsask/deepmedic`

output out_l of layer l and the input in_{l-1} to the previous layer. This choice follows the investigation of [17] that found identity mappings on the residual path to perform best. More formally:

$$in_{l+1}^m = \begin{cases} out_l^m \oplus \hat{in}_{l-1}^m, & \text{if } m \leq M^{l-1} \\ out_l^m, & \text{otherwise} \end{cases} \quad (1)$$

where the superscript m denotes the m^{th} channel and M^l is the number of feature maps in layer l. $\hat{in}_l$ is the input of the previous layer after padding in the (x,y,z) dimensions with reflection in order to match the dimensions of out_l.

3 Experiments

3.1 Data, Pre-processing and Augmentation

Data: We used the training database of BRATS 2015, which includes 220 multi-modal scans of patients with high (HGG) and 54 with low grade glioma (LGG). T1, T1c, T2 and FLAIR sequences are available. The images were registered to a common space and resampled to isotropic $1^3 mm$ resolution by the organisers, with dimensions of each volume 240×240×155. The provided manual segmentation includes four labels: 1) necrotic core (NC), 2) oedema (OE), 3) non-enhancing (NE) and 4) enhancing core (EC). The official evaluation is performed by merging the predicted labels into three sets: whole tumor (all 4 labels), core (1,3,4) and enhancing tumor (4).
Pre-Processing and Augmentation: Each scan was individually normalized by subtracting the mean and dividing by the standard deviation of the intensities within the brain. Training data were augmented via reflection with respect to the sagittal axis.

3.2 Effect of Residual Connections

To evaluate the effect of the residual connections, we performed 5-fold cross validation on the mixed HGG and LGG data. First we reproduced results similar to what was reported in [1] for the original version of DeepMedic. The extension with residual connections gave a modest but consistent improvement over all classes of the task, as shown in Table 1. Important is that performance increases even on small challenging substructures like the necrosis and non-enhancing tumor, which may not be individually evaluated for the challenge but is interesting from an optimization perspective. The improvement seems mainly due to an increase in sensitivity, however at the cost of a lower precision. This is a positive side-effect as in practice it can prove easier to clear false positives in a post-processing step, for instance with a Conditional Random Field as performed in [1], rather than capturing areas previously missed by the CNN.

Table 1: Performance of the original *DeepMedic* (DM) and its extension with residual connections *DMRes*. For consistency with the online evaluation platform, cases that do not present a class in the manual segmentation contribute zero towards the average.

	DSC			Precision			Sensitivity			DSC			
	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	NC	OE	NE	EC
DeepMedic	89.6	75.4	71.8	89.7	84.5	74.3	90.3	73.0	73.0	38.7	78.0	36.7	71.8
DMRes	89.6	76.3	72.4	87.6	82.4	72.5	92.2	75.4	76.3	39.6	78.1	38.1	72.4

3.3 Behaviour of the network with less training data and filters

CNNs have shown promising performance when trained either on the extensive database of BRATS 2015 or on the rather limited of BRATS 2013. However, qualitative differences of the two databases do not allow estimating the influence of the database's size. Additionally, although various architectures were previously suggested, no work has investigated the capacity of the network required for the task. This factor is significant in practice, as it defines the computational resources and inference time required.

We evaluate the behaviour of our network on the tumour segmentation task with respect to the two above factors. To avoid bias towards subjects with more than one acquisitions, only one was used from each subject. Out of the 198 remaining datasets, we randomly chose 40 (29 HGG, 11 LGG) as a validation fold. We then trained the original version of DeepMedic on the remaining 158 datasets, as well as on a reduced number of training data. Finally, versions of the network where all layers had their filters reduced to 50% and 33% percent were trained on the whole training fold.

It can be observed on Table 2 that the network retains most of its performance for the three merged classes of the challenge, even when trained with little data or its capacity is significantly reduced. A more thorough look in the accuracy achieved for the 4 non-merged classes shows that the greatest decline is observed for the challenging necrotic and non-enhancing classes, which however does not influence the segmentation of the overall core as significantly. This shows that both training data and network size do matter for the segmentation of such small and challenging substructures[5].

Table 2: Performance of *DeepMedic* with reduced training data or number of filters.

	DSC			Precision			Sensitivity			DSC			
	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	NC	OE	NE	EC
DeepMedic	91.4	83.1	79.4	89.2	87.7	82.8	94.1	80.8	79.5	50.0	79.6	35.1	79.4
DM(75% data)	91.2	82.5	79.6	89.0	84.4	82.4	93.9	80.4	79.7	45.9	79.0	35.1	79.6
DM(50% data)	91.4	82.6	78.8	91.0	85.3	81.7	92.3	82.3	78.5	44.7	79.2	36.8	78.8
DM(33% data)	90.5	79.7	77.7	90.6	86.5	82.8	91.0	77.1	77.1	45.8	77.9	31.8	77.7
DM(20% data)	89.8	80.5	77.6	91.1	83.9	81.8	89.7	80.5	76.5	41.3	76.9	34.1	77.6
DM(50% filters)	91.4	80.8	79.8	92.2	89.0	82.5	91.3	76.3	80.2	49.0	79.2	29.4	79.9
DM(33% filters)	90.8	81.7	79.5	90.0	91.9	78.2	92.1	76.6	83.0	44.4	79.3	27.9	79.4

4 Discussion and Conclusion

We investigated the effect of residual connections on a recently presented 3D CNN model on the task of brain tumour segmentation, where their incorporation led to a small but consistent improvement. Our work reveals that 3D CNNs gain from such an extension, unlike the observation of [19] who found benefits only for significantly deeper 2D networks. Furthermore, we investigated the behaviour of DeepMedic when trained with less data or when less filters are used. Our experiments show that segmentation accuracy for the whole, core and enhancing tumour is not severely hindered by the two factors. However, they are very important for segmenting challenging, fine substructures such as necrosis and non-enhancing tumour. On the other hand, in applications where segmentation of such substructures is not required, small networks can be a suitable option, thanks to lower computational requirements and shorter inference times (8 versus 35 seconds for the smallest and the original model on an NVIDIA Titan X using cuDNN V5.0). Future investigation will aim at examining network behaviour on other segmentation tasks.

[5] Although these experiments were performed with the original version of the network, we expect the trends to continue after the extension with residual connections.

References

1. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. arXiv preprint arXiv:1603.05959 (2016)
2. Mazzara, G.P., Velthuizen, R.P., Pearlman, J.L., Greenberg, H.M., Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated mri segmentation. International Journal of Radiation Oncology* Biology* Physics **59**(1) (2004) 300–312
3. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Medical image analysis **8**(3) (2004) 275–283
4. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In: MICCAI. Springer (2011) 532–540
5. Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Joint tumor segmentation and dense deformable registration of brain mr images. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012. Springer (2012) 651–658
6. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O., Das, T., Jena, R., Price, S.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In: MICCAI. Springer (2012) 369–376
7. Tustison, N., Wintermark, M., Durst, C., Brian, A.: ANTs and Arboles. in proc of BRATS-MICCAI (2013)
8. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. in proc of BRATS-MICCAI (2014)
9. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in mri images. IEEE transactions on medical imaging **35**(5) (2016) 1240–1251
10. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). Medical Imaging, IEEE Transactions on **34**(10) (2015) 1993–2024
11. Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davatzikos, C.: Glistrboost: Combining multimodal mri segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer (2015) 144–155
12. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. arXiv preprint arXiv:1505.03540 (2015)
13. Lyksborg, M., Puonti, O., Agn, M., Larsen, R.: An ensemble of 2d convolutional neural networks for tumor segmentation. In: Image Analysis. Springer (2015) 201–211
14. Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al.: Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. Medical Image Analysis **35** (2017) 250–269
15. Alansary, A., Kamnitsas, K., Davidson, A., Khlebnikov, R., Rajchl, M., Malamateniou, C., Rutherford, M., Hajnal, J.V., Glocker, B., Rueckert, D., Kainz, B.: Fast fully automatic segmentation of the human placenta from motion corrupted mri. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016 (2016)
16. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
17. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. arXiv preprint arXiv:1603.05027 (2016)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)

Brain Tumor Segmentation from Multi Modal MR images using Stacked Denoising Autoencoders

Varghese Alex and Ganapathy Krishnamurthi

Indian Institute of Technology Madras, Chennai India,
gankrish@iitm.ac.in

Abstract. Accurate automated segmentation of Gliomas is required for evaluating therapy and monitoring tumor growth. For segmentation of Gliomas from Magnetic Resonance Images (MRI), we propose a 5 layer deep Stacked Denoising Autoencoders(SDAE). SDAE's were trained using patches of size 21 x21 extracted from various MRI sequences like T1,T2, FLAIR and T1 post contrast images. The network has 1764 units at the input layer and 3500, 2000,1000, 500, 5 neurons in the subsequent layers. The network was trained, validated and tested exclusively on longitudinal data to quantify the change in lesion as a function of time.

Keywords: Gliomas, Deep Learning, SDAE, MRI

1 Introduction

This paper is in continuation to our submission for BRAT's 2015 [1]. We propose a 5 layer deep Stacked Denoising Autoencoders for segmenting Gliomas from MRI. The network is trained by a combination of unsupervised and supervised technique.

Since the challenge involves determining the technique's ability to the quantifying longitudinal changes, we train, validate the SDAE exclusively on longitudinal data present in the training set.

2 Materials and Methods

Data The training data comprises of 123 images that were acquired only at one time point and 97 images which were acquired from patients over a period of time (longitudinal data). The longitudinal dataset was split into training, validation and test set in the ratio 70:20:10.

The data was preprocessed by matching the histogram of the sequences to a reference sequence [2]. Further each sequence was normalized to have zero mean and unit variance.

II

Table 1. Architecture of the Network

Input Layer	Hidden	Hidden	Hidden	Hidden	Output
1764	3500	2000	1000	500	5

Training of SDAE Training of SDAE comprises an unsupervised phase called pretraining and a supervised phase called fine-tuning. The architecture of the network is given in Table (1)

Patches for pretraining and fine-tuning are extracted from the vicinity of the lesion. Class Balanced mini-batches were used for pretraining,while the network was fine-tuned with unbalanced data. The hyper-parameters i.e. learning rate, optimizer, L1/L2 penalty are given in Table 2.

Table 2. HyperParameters of the Network

	Learning Rate	Learning Rate Decay	L1	L2	optimizer	Noise Level	Dropouts
Pretraining	0.001	0.005	-	0.0001	RmsProp	0.25	-
Fine-tuning	0.005	0.01	-	0.00005	SGD	-	0.25

3 Results

The network was tested on 3 longitudinal patients(9 image volumes) and as a preliminary result we achieve a mean whole tumor,tumor core and active tumor dice score of 0.84, 0.71 and 0.81 respectively. Further statistics of the network is given in Table (3)

Table 3. Performance of Network on Test Data

	Min	25	50	75	Max	Mean	Std
Whole Tumor	0.67	0.83	0.85	0.90	0.93	0.84	0.08
Tumor Core	0.46	0.61	0.77	0.80	0.89	0.71	0.14
Active Tumor	0.56	0.78	0.84	0.87	0.90	0.81	0.11

3.1 Conclusion

We propose a fully automated technique based on SDAE's for segmentation of Gliomas from MR images. The network was trained and validated exclusively on longitudinal data so as to learn longitudinal changes in lesion.

On test data the network achieves mean whole tumor, tumor core and active tumor dice score of 0.84,0.71 and 0.81 respectively.

III

References

1. Vaidhya et al.: "Multi-modal Brain Tumor Segmentation Using Stacked Denoising Autoencoders". Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 181–194 (2016)
2. Nyul et. al. , "New Variants of a Method of MRI Scale Standardization", IEEE Transactions on Medical Imaging, 19(2):143-150, 2000

Brain Tumor Segmentation Using Deep Convolutional Neural Network

Tseng Kuan Lun and Winston Hsu

National Taiwan University

Abstract. We present a fully-automatic brain tumor segmentation method by utilizing Convolutional Neural Network (CNN). CNN-based methods have been applied in brain tumor segmentation. However, most of them are patch-wise learning models, which ignore the contextual information within the whole image region. Moreover, the problem of label imbalance leads the CNN models often converge to the certain labels. We tackle these problems by leveraging global-based CNN methods (e.g., FCN or SegNet) and incorporating multi-modalities of MRI data. We adopt a re-weighting scheme on the CNN loss layer to alleviate the problem of label imbalance. Preliminary results on BRATS-2015 dataset demonstrate that global-based CNN models with re-weighting outperform the prior patch-wise learning methods.

Keywords: Brain tumor segmentation, deep neural networks

1 Introduction

End-to-end convolution neural networks (CNN) have drawn increasing attention recently, e.g., in object detection, segmentation, and classification. To our knowledge, most of CNN-based methods for BRATS challenge are patch-wise models [3] [6]. However, they often ignore the contextual information in the whole image region, which is considered informative for brain tumor segmentation, and require further post-processing steps to refine the results. Mohammad Havaei et al purpose a two-path CNN model [3] attempting to overcome this issue. They use different crop sizes and different receptive fields to get the global information. Motivated by the success of global-based CNN methods: Fully Convolution Networks (FCN) [5] and SegNet [1] for generic image segmentation, we leverage them to obtain the dense labels for brain tumor segmentation. Global-based methods are faster than patch-wise methods and directly capture the tumor structure and position in the brain. In our preliminarily experiments on BRATS-2015 dataset, they outperforms the conventional patch-based methods.

There are four labels to describe the types of tissues and four different modalities for MRI data, e.g., T1 (spin- lattice relaxation), T1- contrasted (T1C), T2 (spin-spin relaxation) and fluid attenuation inversion recovery (FLAIR) pulse sequences. In our method, we take these modalities as different channels, i.e., four-channel image slices, and use whole images as inputs to train end-to-end

CNN models. For each brain image, we generate the input data by using depth slicing. For example, if the dimension of the original data is $4 \times 240 \times 240 \times 155$, there are 155 image slices after depth slicing.

2 Method Overview

We utilize two end-to-end global models and a patch-wise model to segment the brain images in BRATS-2015 dataset. The global models are Fully-Connected Neural Network (FCN) [5] and SegNet [1], which show state-of-the-art performance on semantic image segmentation. Patch-wise model performs the prediction at each local region while global-based models take full-size images as inputs and predict the results by exploiting the full image information.

2.1 Patch-Wise Model

For patch-wise model, we extract each local region by cropping an image patch with 30×30, and predicting the label (center pixel) of each patch. Our implementation of patch-wise learning model uses five convolution layers followed by one ReLU layer and two fully connected layers. We only use one pooling layer after the first convolution layer. The parameters in this architecture is relatively small compared to standard classification net such as AlexNet [4] or VGGNet [7]. The reason of using this architecture is that the size of BRATS dataset is relatively small and large parameters often leads to overfitting.

2.2 Fully Convolutional Network

The core idea of fully convolutional networks [5] is that it changes the fully connected layers of VGG-16 into convolution layers. Thus, enable a classification net to output a dense heat map and preserve the spatial information. Due to the large size of FCN models with VGG structure, we train FCN using AlexNet without the skip layers.

2.3 SegNet: Encoder-Decoder Architecture

SegNet [1] uses Encoder-Decoder architecture similar to FCN. The encoder network is identical to VGG-16 with the fully-connected layer removed. The decoder network up-samples the heat map according to the indexes used in the pooling layers of the encoder network, which is different from the up-sampling strategy used in FCN. The high-level semantic information in the upper layers are automatically preserved in each up-sampling step. Since BRATS dataset is relatively small, SegNet is more suitable (i.e., less parameters) compared to FCN.

3 Model Learning

For training FCN and SegNet, we use depth slicing images and the labels from ground truth data as inputs. To deal with the data imbalance problem (e.g., the healthy tissue is about 98 percent of the data, and FCN model often converges to healthy tissues), we adopt median frequency balancing [2], where the weight assigned to a class in the loss function is the ratio of the median of class frequencies on the entire training set divided by the class frequency. To distinguish the black padding (background) from the normal tissue (they are with the same label 0), we relabel the black padding voxels to label 5. We crop a large numbers of 30×30 image patches for training the patch-wise model and ensure the numbers of images for different labels are consistent. Since the input data has four channels, we do not use a pre-trained model such as AlexNet and VGGNet.

4 Result

Table 1 shows the results of the three different models evaluated at BRATS 2015 online judge system. The judge system uses three evaluation tasks. Complete score represent whole tumor area therefore it evaluates all labels 1,2,3,4. (0 for normal tissue, 1 for edema, 2 for non-enhancing core, 3 for necrotic core, 4 for enhancing core). Core score represent only tumor core region, it measures label 1,3,4. The last one is enhancing score, which represent the active tumor region (only containing the enhancing core(label 4) structures that are unique to high-grade cases). Enhancing score is the hardest to optimize because the few training samples of label 4. There are three kinds of evaluation criteria, Dice, Positive Predicted Value and Sensitivity.

$$Dice = \frac{P_1 \bigcap T_1}{(P_1 + T_1)/2}$$

$$PPV = \frac{P_1 \bigcap T_1}{P_1}$$

$$Sensitivity = \frac{P_1 \bigcap T_1}{T_1}$$

Which T stands for ground truth label and P stands for predicted result. T_1 is the true lesion area and P_1 is the subset of voxels predicted as positives for the tumor region.

From the results, SegNet outperforms the other models in different label sets (i.g., complete, core and enhancing), especially for label 3, 4. compared to path-wise model, validating the feasibility of leveraging global contextual information for brain tumor segmentation.

Table 1. Evaluation results at BRATS 2015 online judge system.

Dice	complete	core	enhancing
Patch-Wise	0.74	0.64	0.60
FCN	0.71	0.58	0.47
SegNet	**0.75**	**0.77**	**0.76**

PPV	complete	core	enhancing
Patch-Wise	0.65	0.59	0.58
FCN	0.66	0.56	0.55
SegNet	**0.68**	**0.71**	**0.72**

Sensitivity	complete	core	enhancing
Patch-Wise	**0.92**	0.78	**0.72**
FCN	0.82	0.69	0.50
SegNet	0.87	**0.85**	0.64

5 Conclusion

We investigate three different models for brain tumor segmentation and demonstrate that global-based CNN methods with re-weighting strategy outperform patch-wise learning method. For the future work, we seek further improvements by incorporating temporal information (e.g., LSTM) to model the relationships between different depths.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arxiv:1511.00561 (2015)
2. D.Eigen, R.Fergus: Predictingdepth,surfacenormalsandsemantic labels with a common multi-scale convolutional architecture. arXiv preprint arXiv:1411.4734 (2014)
3. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. Volume 35, January 2017, Pages 18-31. (2017)
4. Krizhevsky, A., Sutskever, I., , Hinton, G.: Imagenet classification with deep convolutional neural networks. NIPS (2012)
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. arXiv preprint arXiv:1411.4038 (2014)
6. Pereira, S., Pinto, A., Alves, V., Silva, C.: Brain tumor segmentation using convolutional neural networks in mri images. IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1240–1251 (2016)
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014)

Brain Tumor Segmentation with Optimized Random Forest

László Lefkovits[1], Szidónia Lefkovits[2], and László Szilágyi[1]

[1] Department of Electrical Engineering
Sapientia University of Transylvania Tîrgu-Mureș, Romania
lefkolaci@ms.sapientia.ro
[2] Department of Computer Science
,,Petru Maior" University Tîrgu-Mureș, Romania

Abstract. In this article we propose and tune a discriminative model based on Random Forest (RF) to accomplish brain tumor segmentation in multimodal MR images. The objective of tuning is to establish the optimal parameter values and the most significant constraints of the discriminative model. During the building of the RF classifier, the algorithm evaluates the importance of variables, the proximities between data instances and the generalized error. These three properties of RF can be applied to optimize the segmentation framework. At the beginning the RF is tuned for variable importance evaluation, and after that it is used to optimize the segmentation framework. The framework was tested on unseen test images from BRATS. The results obtained are similar to the best ones presented in previous BRATS Challenges.

Keywords: random forest; feature selection; variable importance; statistical pattern recognition; MRI segmentation

1 Introduction

In this paper we propose a discriminative model for brain tumor segmentation in multimodal MRI. The main part of the model is the evaluation and selection of low level image features for segmentation task. Another goal is the optimization of the RF classifier. This is achieved in two phases: first, we start with the optimization of RF considering variable importance evaluation. Secondly, the RF structure is optimized in order to improve segmentation performances. Many discriminative segmentation models have been proposed and tested in the last four BRATS Challenges (2012-2015) [4]. The selection of features used was based on the intuition and experience of the authors. The exact definition and usage of features applied remains their secret [6, 7, 10, 11]. Usually, the systems work with a large amount of features that have hardly any theoretical or practical analysis behind their utility.

2 The Discriminative Model Proposed

The discriminative model proposed is similar to previously used models (figure 1), but in this article we would like to emphasize certain interesting aspects which have an important contribution to the performances reached.

The performances of a segmentation model built on a discriminative function are mainly determined by three important issues: the quality of the annotated image database, the classification algorithm applied and the feature set used.

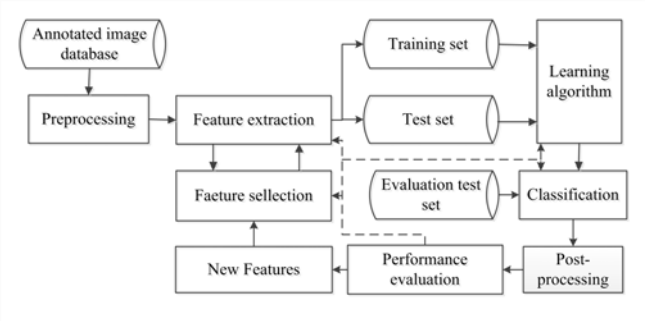


Fig. 1: Proposed Discriminative Model

2.1 Database

The most important and probably, the only image database for brain tumor segmentation was created during the BRATS Challenges (2012-2015), thanks to B. Menze and A. Jakab [9]. In our work we consider the same three classes used for performance evaluation in BRATS Challenges. These classes are: whole tumor - WT (including all four tumor structures), tumor core - TC (including all tumor structures except edema) and active tumor - AT (only the enhancing core).

The principle of statistical pattern recognition is the assignment of certain features to every voxel. The resulting database increases linearly with each feature added; thus, if one decides to use a set of about 1000 features, a total of 84 TB of information has to be processed. The more data can be processed and included in the training phase, the better the performance of the classifier obtained. In our work we solved the problem of this huge and unmanageable dataset in two ways: 1. by reducing the irrelevant and redundant features; 2. by eliminating similar cases.

2.2 Preprocessing

In our work we have analyzed three important artifacts: inhomogeneity correction, noise filtering and intensity standardization. For inhomogeneity reduction in MR images, we have applied the N4 filter implemented in the ITK package [1].

For noise reduction we used the anisotropic filtering from the same package. Intensity normalization was done by histogram linear transformation in such a way that the first and third quartiles have predefined values.

2.3 Feature Extraction

For each voxel we defined many low-level image characteristics that describe the intensities in the neighborhood of the voxel studied. In our application we have used the following features: first order operators (mean, standard deviation, max, min, median, Sobel, gradient); higher order operators (laplacian, difference of gaussian, entropy, curvatures, kurtosis, skewness); texture features (Gabor filter); spatial context features (symmetry, projection, neighborhoods).
We extracted 240 image features of each modality and we obtained a feature vector with 960 elements. All of these features are defined in the Weka Segmentation plugin found in the Fiji package [2].

2.4 Random Forest

A powerful algorithm for segmentation purposes, Random Forest (RF) [5] has five important characteristics that make it applicable for segmentation tasks: it manages large databases easily; it handles thousands of variables; it estimates the variable importance used in classification; it is able to balance the error in unbalanced datasets; and it produces an internal unbiased estimator of generalized error. All of these characteristics are very important and widely used in our optimization framework.

The RF classifier is built on two random processes: the random build of the bootstrap set and the random feature selection in each node. The creation of RF is based on two sets: the bootstrap set, containing the instances for building a tree and the OOB set (out-of-bag set), containing test instances not included in the bootstrap set.

2.5 Feature Selection

The main part of the model is the evaluation and selection of low-level image features for segmentation task. In the field of statistical pattern recognition, the selection of such features is a challenging task. In order to create a well-functioning discriminative model, we have to select the features relevant for our application and eliminate the useless ones. For this purpose we use the variable importance evaluation provided by RF. In our algorithm we used only the permuted variable importance, also called mean in accuracy. In order to deal with many instances, each consisting of a large number of features, we propose to use our feature selection algorithm, presented in detail in [8]. The main idea of the algorithm is to evaluate the variable importance several times on a randomly chosen part of training set. It eliminates the most unimportant 20-50% of variables at each run. In our experiment, after each reduction step, the RF classifier

was trained and evaluated in order to determine the segmentation performances. Furthermore, the elimination of variables depends on the decrease in segmentation performances. The proportion of reduction is empirical; it depends on the number of attributes used and the performances obtained.

We extracted 240 low level features from every 4 image modality. These 960 features require a total of 10 GB of memory for one 3D brain image. In order to provide an appropriate training set, we had to use the information from at least 20 brain images. Thus, the whole training database is approximately 200 GB in size, which is hardly manageable. There are two ways to reduce this size: reducing the number of instances and reducing the number of features. After this cut off in the number of instances, the sampled training database size is reduced to 0.5 GB. This size is manageable by our system (Intel(R) Core(TM) i7-2600k CPU@3.40 GHz, 16 GB RAM) by using the RF implementation from the R package provided by CRAN [3].

2.6 RF optimization

The RF algorithm generates, during the built of the classifier, an internal error, namely the OOB error. This error is the mean value of classification error of each tree on its own OOB sets. The main task is to find some correlation between the OOB error and the Dice coeficient of segmentation. There are three parameters, that have to be tuned in RF optimization: the number of trees K_{trees}, the number of features m_{tries} and the number of nodes used in each tree T_{nodes}.

2.7 Post-processing

In post-processing we took into consideration the spatial relation between the different tumor classes. In this way, we were able to eliminate many small volumes that had been interpreted as false detections.

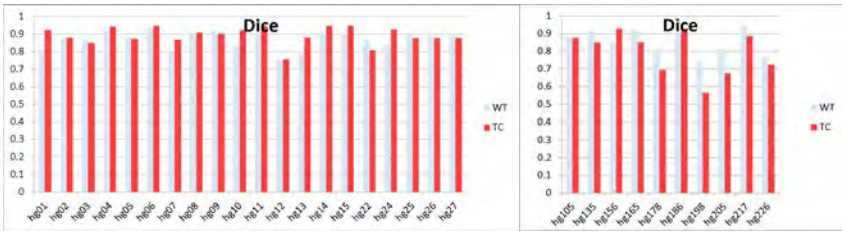
3 Results and Conclusion

The training set is made up of 250000 ($< 1\%$) randomly sampled instances from a total of 30 million instances, representing 20 HG images (the BRATS 2013 training set). Our classifier based on the optimization framework proposed, is composed of $K_{trees} = 120$ trees, each with a size of $T_{size} = 4096$ nodes and $m_{tries} = 9$ randomly chosen features from the reduced feature set ($M = 120$ features). The classification results obtained on 20 HG images from BRATS 2013 and 10 HG images from BRATS 2015 are given in fig. 2a and fig. 2b, respectively. Table 1 shows the Dice coefficients obtained on 50 HG unseen test images using the online evaluation website [4]. These performances are comparable with previously reported results.

We are working on further optimization in order to improve segmentation performances. The goal is to build an optimized training set containing all tumor appearances from the BRATS 2015 datasets. This can be done by adding

Table 1: Compared Dice indexes

HG	Our test classif.	Brats 2012 [9]	Brats 2013 [9]
WT	82 [%]	63-78 [%]	71-87 [%]
TC	70 [%]	24-37 [%]	66-78 [%]



(a) 20 HG images BRATS 2013

(b) 10 HG images BRATS 2015

Fig. 2: Segmentation results

new instances to the training set and also by measuring their proximity to the existing set.

Furthermore, the segmentation framework obtained can be enlarged by new low-level features. The importance of new features could be evaluated and compared to the current set of features. In this way we are able to create a better feature set for tumor segmentation. Besides, additional post-processing may lead to supplementary improvement in segmentation performances.

References

1. ITK. http://www.itk.org/
2. Trainable segmentation. http://imagej.net/Scripting_the_Trainable_Segmentation
3. The R project for statistical computing. https://www.r-project.org/
4. The SICAS Medical Image Repository. https://www.smir.ch/BRATS/Start2015
5. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
6. Geremia, E., Menze, B.H., Ayache, N.: Spatial decision forests for glioma segmentation in multi-channel mr images. In: MICCAI-BRATS Challenge (2012)
7. Goetz, M., Weber, C., Bloecher, J., Stieltjes, B., et al.: Extremely randomized trees based brain tumor segmentation MICCAI-BRATS Challenge (2014)
8. Lefkovits, L., Lefkovits, S., Vaida, M.F.: An Optimized Segmentation Framework Based on Random Forest. Studies in Informatics and Control (2016), under review
9. Menze, B.H., Jakab,A., Bauer,S. et al.: The multimodal brain tumor image segmentation benchmark(BRATS).IEEE Tr.on Medical Imaging 34(10),1993–2024, 2015
10. Reza, S., Iftekharuddin, K.M.: Multi-class abnormal brain tissue segmentation using texture features. In: MICCAI-BRATS Challenge (2013)
11. Zikic, D., Glocker, B., Konukoglu, E., et al.: Context-sensitive classification forests for segmentation of brain tumor tissues. In: MICCAI-BRATS Challenge (2012)

Segmentation of Brain Tumors via Cascades of Lifted Decision Forests

Loic Le Folgoc, Aditya V. Nori, Javier Alvarez-Valle, Richard Lowe, and Antonio Criminisi

Microsoft Research, Cambridge, UK

Abstract. We revisit the well-known decision forest (DF) framework for automatic segmentation of brain tumors in multi-channel MR images. Our approach builds upon previous work using the following ideas: 1) *semantic auto-context*: it exploits label semantics to decompose the segmentation task into a sequence of segmentation subtasks, and 2) *fast lightweight training*: it couples simple generic intensity-based features, together with a novel node-splitting criterion that effectively introduces node-level cross-validation in DFs. This results in tremendous gains in terms of generalization capabilities as well as computational cost, both at training and test time. The training time per tree is $\sim$ 10 seconds, while the test time per image is $\sim$ 30 seconds. With training on 10 images from the BRATS 2013 dataset, the proposed model achieves at test time a DICE score of respectively 79%, 67% and 70% for the whole tumor, tumor core and enhancing parts over the rest of the BRATS 2015 dataset.

1 Introduction

The past few years have witnessed a vast body of machine learning (ML) techniques for the automatic segmentation of medical images. Decision forests [1], and more recently, deep neural networks [2] have yielded state-of-the-art results at MICCAI Brain Tumor Segmentation challenges [3][4]. We build upon the existing framework of decision forests (DFs). Our proposed approach addresses the issue of overfitting, one of the most glaring limitations of DFs. It is comparable with state-of-the-art techniques accuracy-wise, despite using generic features and only a fraction of their computational resources. The intuition for this is as follows:

Semantic auto-context. We exploit label semantics to hierarchically decompose a difficult medical image segmentation task into a sequence of simpler subtasks. Each subtask is solved with a DF layer. Layers are trained one at a time and cascaded: the output of upstream layers augments the input of subsequent layers to provide semantic context. In addition, the region of interest (ROI) is progressively refined to boost the discriminative power of downstream layers.

2

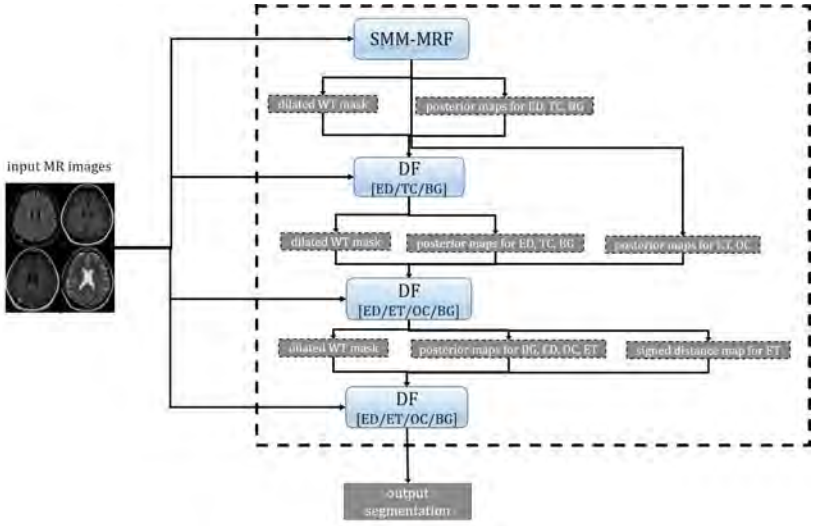


Fig. 1. The cascades of lifted DFs model.

Fast lightweight learning. Annotated data is often scarce and costly in medical imaging; therefore increasing the learning rate of ML algorithms is crucial. We propose a node-splitting criterion that effectively introduces node-level cross-validation in the DF framework at training time, at no additional computational cost. This greatly improves the generalization capability of DFs, and we call them *lifted DFs*. By lifting the discriminative power of each node, we are able to outperform classical DFs using shallower trees with fewer nodes. Coupled with bagging [5] (randomized subsets of training data), this results in a tremendous decrease in training time—each tree in the forest can be trained within seconds. We leveraged this benefit of fast experimentation to design features that have good discriminative power and are computationally efficient.

2 Overview

Given 4 distinct MR modalities (T1-Gadolinium, T1, T2, FLAIR) as input, our model (shown in Figure 1) performs a 4-way classification of voxels into the following mutually exclusive classes: enhancing tumor part (ET), other parts of the tumor core (OC), the necrotic core (NC) and non-enhancing parts (NE), edema (ED) and background (BG). The enhancing tumor together with the other parts of the tumor core form the tumor core ($TC = ET \cup OC$), which together with the edema is referred to as the whole tumor ($WT = TC \cup ED$).

As shown in Figure 1, we proceed in a cascade of steps, each of which either addresses a coarser subtask w.r.t. the hierarchy of labels (*e.g.*, locating the region of interest for WT), and/or provides approximate posterior probability maps fed as semantic spatial context to the subsequent layer(s).

An SMM-MRF layer is used to locate the region of interest (ROI) for the whole tumor. The likelihood for each of the five ground truth classes is modelled using a Student Mixture (SMM) with spatially varying (BG) or fixed (other classes) proportions [6], as a suitably modified variant of [7]. An MRF prior is assumed for the WT and BG, and the model parameters are learnt via Variational Bayesian inference. The output WT segmentation map is dilated (1.5cm): this defines a mask over which the next layer operates.
A first DF layer performs a 3-way classification: ED vs. TC vs. BG. The input channels for this layer are the four original MR intensity channels, posterior maps for ED, TC and BG provided by the SMM-MRF model as well as the spatially varying mixture proportions for the background SMM. The ROI is further refined by dilation of the output WT segmentation (5mm).
A second DF layer performs a 4-way classification: BG vs. ED vs. OC vs. ET. It takes as input the 4 MR intensity channels, 3 posterior probability maps provided by the previous layer (ED, TC and BG) as well as the ET and OC posterior maps derived from the SMM-based model. The ROI is further refined by dilation of the output WT segmentation (3mm).
A final DF layer is trained to further refine the classification. It takes as input the 4 MR intensity channels, and the 4 probability channels (BG, ED, OC, ET) output by the previous layer. In addition, the signed distance map to the region enclosed by the enhancing tumor (ET) is computed and passed as a channel to help disambiguate between ED and OC.

The same pipeline is followed at training and test time. The benefit of restricting the training of successive layers to a refined ROI provided by the previous layer is to improve training accuracy by removing unrelated background clutter and focusing on relevant training examples. This can be interpreted as a very practical form of boosting. The drawback is that it is possible to miss parts of the actual tumor for the downstream cascaded layers.

3 Technical Details

DF training. Let x be a point to be classified, and $I = \{I_j\}$ the set of input channels. Routing of points through nodes happens via binary predicate evaluations $[f(x, I) < \theta]$, where $f(x, I)$ is a feature. The candidate features for tree node splitting are the same for all layers, and are selected as follows: at training time, the optimal feature f_i and threshold θ_i are typically chosen to be the ones that minimise the entropy over a random set of candidate features $f_i, i = 1 \dots F$, and a random training subset of examples D_T. This is equivalent to maximum likelihood: $(f_i^*, \theta_i^*) = \text{argmax}_i \, \text{argmax}_{\theta_i} \, p(D_T | f_i, \theta_i)$.

We make the following modification to this selection process: the optimal threshold per feature is still selected based on entropy, but the optimal feature is selected using cross-entropy on a distinct validation subset of examples D_V. This is equivalent to node-level cross validation: $(f_i^*, \theta_i^*) = \text{argmax}_i \, p(D_V | f_i, \theta_i^*)$ s.t. $\theta_i^* = \text{argmax}_{\theta_i} \, p(D_T | f_i, \theta_i)$.

Fast context-aware features. The features $f(x, I)$ belong to one of 5 intensity-based, parametrized feature types. The first type simply probes a channel intensity at the voxel of interest: $f^1(x, I)_j = I_j(x)$. The 4 remaining types are fast generic context-aware features. They operate on smoothed versions I_j^k of the input channels among preset scales $\{s_k\}$. They consist of scale-space intensity, intensity difference, "spherical difference" and "spherical likeness" feature types which are defined as follows:

$$
\begin{aligned}
f^2(x, I)_{j,k} &= I_j^k(x)\,, \\
f^3(x, I)_{j_1,j_2,k_1,k_2,v} &= I_{j_2}^{k_2}(x+v) - I_{j_1}^{k_1}(x)\,, \\
f^4(x, I)_{j_1,j_2,k_1,k_2,r} &= \mathrm{Median}_{n=1}^{12} \left(I_{j_2}^{k_2}(x + r \cdot \phi_n) - I_{j_1}^{k_1}(x) \right)\,, \\
f^5(x, I)_{j_1,j_2,k_1,k_2,r} &= \mathrm{Median}_{n=1}^{12} \left| I_{j_2}^{k_2}(x + r \cdot \phi_n) - I_{j_1}^{k_1}(x) \right|\,.
\end{aligned}
$$

where ϕ_n, $n = 1 \ldots 12$, are the vertices of a regular icosahedron.

4 Results

The proposed approach is implemented within the DF framework described in [1]. All experiments were performed on a 3.6GHz Intel Xeon processor system with 16GB RAM running Microsoft Windows 10. The model is trained on a set of 10 images from the BRATS 2013 dataset, tested on 10 images from the BRATS 2013 dataset and 200 images in the BRATS 2015 dataset. Our results are summarized in Table 1.

It is interesting to note that the results on the BRATS 2013 dataset are quantitatively on par with the Decision Forest based winning method [3, 8] with orders of magnitude of reduction in computational time and fewer training images. Similarly, the results are also promising on the whole BRATS 2015 dataset, thereby demonstrating the generalization capability of the proposed method.

In a second experiment, we train the model on 1/5 of the whole BRATS 2015 HGG dataset (44 images) and evaluate it on the remaining 4/5th, achieving DICE scores of 82%, 73% and 75% resp. for WT, TC and ET.

	Dice(%)		
Dataset	WT	TC	ET
BRATS 2013	88	76	79
BRATS 2015	79	67	70

Table 1. Results over the BRATS 2013 and 2015 datasets when trained on 10 images. Each layer uses 100 trees of maximum depth 8. Training time per tree $\sim$ 10 seconds. Total training time for the cascade of DFs $\sim$ 1 hour, training trees sequentially. Test time per image $\sim$ 30 seconds.

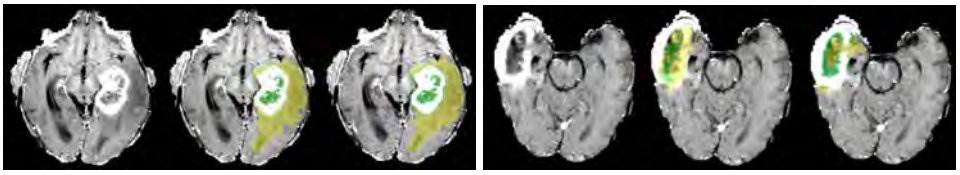


Fig. 2. Example results. For each of the 2 displayed subjects: 3 images resp. showing: the T1-Gadolinium channel (left), the ground truth (middle) and the prediction (right). The leftmost subject is among the quantitatively better results and the rightmost among the poorer. Common mistakes include oversegmentation (false positives, disconnected components) and in some cases confusion between ED and NC.

5 Conclusion

We have presented preliminary results for a novel framework using cascades of lifted DFs that results in good generalization with low computational resources. Next steps include a more comprehensive evaluation over the BRATS2015 dataset—in addition to reporting more detailed statistics, we will also measure the robustness of our approach to the choice of training sets.

References

1. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O., Das, T., Jena, R., Price, S.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2012) 369–376
2. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. arXiv preprint arXiv:1603.05959 (2016)
3. Tustison, N., Wintermark, M., Durst, C., Avants, B.: Ants andarboles. Multimodal Brain Tumor Segmentation (2013) 47
4. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Deep convolutional neural networks for the segmentation of gliomas in multi-sequence mri. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer (2015) 131–143
5. Breiman, L.: Bagging predictors. Machine learning **24**(2) (1996) 123–140
6. Archambeau, C., Verleysen, M.: Robust bayesian clustering. Neural Networks **20**(1) (2007) 129–138
7. Cordier, N., Delingette, H., Ayache, N.: A patch-based approach for the segmentation of pathologies: Application to glioma labelling. IEEE Transactions on Medical Imaging **35**(4) (2015)
8. Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging **34**(10) (2015) 1993–2024

Nabla-net: a deep dag-like convolutional architecture for biomedical image segmentation: application to high- and low-grade glioma segmentation

Richard McKinley, Roland Wiest, Mauricio Reyes

Abstract

Biomedical image segmentation requires both voxel-level information and global context. We present a deep convolutional architecture which combines a fully-convolutional network for local features and an encoder-decoder network in which convolutional layers and maxpooling compute high-level features, which are then upsampled to the resoultion of the initial image usiong further convolutional layers and tied unpooling. The networks are initially trained on data from patients treated at the Inselspital, Bern, and subsequently applied to cases from the MSSEG MICCAI grand challenge.

Introduction

Medical image segmentation is a fundamental problem in biomedical image computing. Manual segmentation of medical images is both time-consuming and prone to substantial inter-rater error: automated techniques offer the potential for robust, repeatable and fast image segmentation. In the past few years, neural networks have returned to the fore as the most promising technique for supervised machine learning.[1] In particular, convolutional neural networks have dominated the field of image recognition. More recently, techniques for image recognition have been reworked to object location and segmentation. One recurring theme is that segmentation requires a combination of low-level and high-level features, with several techniques and architectures having been suggested for upscaling and incorporating high-level with low-level features. [2]–[5] In this paper, we describe an architecture, called nabla net, for image segmentation, with application in the medical image segmentation domain. Nabla net is a dag-like deep neural network architecture, combining a fully-convolutional pathway learning low-level features and an encoder-decoder network learning high-level features. We describe the general features of nabla-net, and its application to the segmentation gliomas.

Method

The fundamental basis of nabla net is a deep encoder/decoder network. Such a network comprises a series of encoder layers, each of which reduces feature maps with higher spatial dimensions to feature maps with lower spatial dimension, followed by a series of decoder layers, which expand features with low spatial dimensions to features with high spatial dimensions. The output of a pure encoder-decoder network provides a good localisation of lesion tissue, but is unable to provide crisp boundaries: for that reason, the final prediction of a nabla net is produced by combining the layers arising from the first encoder layer with the output of the final decoder layer.

We make use of maxpooling [6], with non-overlapping pool size 2. To scale up the feature maps, we use "tied unpooling" [7], [8], in which feature maps are upscaled by filling only the positions of the maximums of the corresponding maxpooling layer. We use make use of batch normalization [9] to accelerate the learning process. The architecture of the nabla net makes no use of fully-connected layers, and as such it can be applied to any image with dimensions divisible by eight, this being enforced by the number of maxpooling steps.

For training data, we took the 244 TCIA cases from the BRATS dataset: these cases have been labelled by a consensus of previously successful segmentation methods: the label set is thus rather noisy, and learning from them is a good test of a method's ability to learn from noisy, unreliable labels.
We segment the volume by complete axial slices, using data from five consecutive axial slices, with such a set of slices being included as training data if the middle slice contained voxels within the brain mask. Ground truth for such a set of slices was given by the lesion mask of the central slice: thus, the lesion mask of a slice was predicted from the slice itself and the four slices surrounding it.

The nabla net is built from encode and decode layers. Each encode layer has the following structure:

Layer	Name	Dimension of output
1	Input	(p, q, r)
2	Zero padding	(64, q+2, r+2)
3	3 by 3 convolutional	(64, q, r)
4	Batch Normalization	(64, q, r)
5	ReLu	(64, q,r)
6	Zero padding	(64, q+2, r+2)
7	3 by 3 convolutional	(64, q, r)
8	Batch Normalization	(64, q, r)
9	ReLu	(64, q,r)

Each decode layer has the following structure

Layer	Name	Dimension of output
1	Input	(p, q, r)
2	Zero padding	(64, q+2, r+2)
3	3 by 3 convolutional	(64, q, r)
4	Batch Normalization	(64, q, r)
5	Zero padding	(64, q+2, r+2)
6	3 by 3 convolutional	(64, q, r)
7	Batch Normalization	(64, q, r)

The initial portion of the network consists of four small encode layers (with 32, rather than 64 feature maps), each working on a separate modality from { T1, T1c, T2, FLAIR}. The output of these four encode layers is then fed into the encoder-decoder network:

Layer	Name	Dimension of output	Comments
1	Input	(4 * 32, n*8, m*8)	Five slices, dimensions (n*8, m*8)
2	Encode 1	(64, n*8, m*8)	
3	Maxpool 1	(64, n*4, m*4)	
4	Encode 2	(64, n*4, m*4)	
5	Maxpool 2	(64, n*2, m*2)	
6	Encode 3	(64, n*2, m*2)	
7	Maxpool 3	(64, n, m)	
8	Encode 4	(64,n,m)	
9	Decode 4	(64,n,m)	
10	Unpool 3	(64, n*2, m*2)	Tied to Maxpool 3
11	Decode 3	(64, n*2, m*2)	
12	Unpool 2	(64, n*4, m*4)	Tied to Maxpool 2
13	Decode 2	(64, n*4, m*4)	
14	Unpool 1	(64, n*8, m*8)	Tied to Maxpool 1
15	Merge	(128 , n*8, m*8)	Concatenate the outputs of layers 2 and 14
16	Encode Final	(64, n*8, m*8)	
17	1*1 Convolutional	(4,n*8, m*8)	
18	Sigmoid	(4, n*8, m*8)	Loss = binary crossentropy

The output of the model has four channels, each corresponding to one of the four classes { Whole Tumour, Gross Tumour, Contrast enhancing, Necrosis}. The network was built using the Keras framework, and trained using Theano as a backend, using Adadelta. The pipeline for lesion segmentation was as follows: apply the network to obtain heatmaps for the whole tumour, gross tumour, contrast-enhancing tumour and necrosis. The final lesion maps are extracted from the heatmaps using a random walk segmentation, as implemented in the

python scikit-image package, for each of the four classes . All voxels outside the whole tumour are labelled as non-tumour: voxels inside the gross tumour but not inside the contrast-enhancing or necrosis are labelled as noncontrast-enhancing, and voxels outside the gross tumour but in the whole tumour are labelled as edema.
When applied to the 30 BRATS 2012 training cases (which were not used for training), the method achieved a mean Dice score of 0.87 for the whole tumour, 0.69 for the tumour core, and 0.56 for contrast-enhancing tumour. A subsequent retraining on these cases will be performed before the method is applied to the BRATS 2016 test cases.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proceedings of the iEEE computer society conference on computer vision and pattern recognition*, 2015, vols. 07-12-June-2015, pp. 447–456.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the iEEE computer society conference on computer vision and pattern recognition*, 2015, vols. 07-12-June-2015, pp. 3431–3440.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015.

[5] T. Brosch, L. Tang, Y. Yoo, D. Li, A. Traboulsee, and R. Tam, "Deep 3D Convolutional Encoder Networks with Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation," *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, p. 1, 2016.

[6] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *Proceedings of the iEEE international conference on computer vision*, 2009, pp. 2146–2153.

[7] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proceedings of the iEEE international conference on computer vision*, 2011, pp. 2018–2025.

[8] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *Cvpr 2015*, p. 5, 2015.

[9] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv:1502.03167*, pp. 1–11, 2015.

CRF-based Brain Tumor Segmentation: Alleviating the Shrinking Bias

Raphael Meier[1], Urspeter Knecht[2], Roland Wiest[2], and Mauricio Reyes[1]

[1] Institute for Surgical Technologies and Biomechanics, University of Bern, Switzerland
[2] Support Center for Advanced Neuroimaging – Institute for Diagnostic and Interventional Neuroradiology, University Hospital and University of Bern, Switzerland
raphael.meier@istb.unibe.ch

Abstract. This paper extends a previously published brain tumor segmentation method with a dense Conditional Random Field (CRF). Dense CRFs can overcome the shrinking bias inherent to many grid-structured CRFs. The proposed segmentation method is evaluated using data of the MICCAI BRATS 2013 & 2015 data sets (up to 110 patient cases for testing) and compared to a baseline method using a grid-structured CRF. Improved segmentation performance was observed with respect to grid-structured CRFs.

1 Introduction

Image segmentation plays a pivotal role in the analysis of medical imaging data. Information extracted via segmentation (e.g. volume/position) can be used for diagnosis, treatment planning and monitoring. Graphical models offer a way to describe structure in medical images and to embed it in a probabilistic framework. Markov or Conditional Random Fields (CRFs) have become one of the most widely-used graphical models in image understanding [10]. The most popular CRFs correspond to a grid, where voxels are represented by nodes in a graph. Smoothness of the final segmentation is induced via pairwise potentials defined for neighboring nodes. The use of a Potts prior in grid-structured pairwise CRFs is known to result in excessive smoothing (also known as *shrinking bias*) of the object boundary. Dense CRFs establish pairwise potentials between all nodes in a graph allowing for long-range interactions between voxels. As a consequence, the inherent shrinking bias of grid-structured pairwise CRFs is alleviated allowing for more detailed segmentation results. Krähenbühl et al. proposed an efficient inference [5] and learning algorithm [4] for dense CRFs, making their application in medical image volumes feasible. In this paper, we extend our previously published methodology [6] with a dense CRF and investigate its impact on the segmentation of tumor subcompartments.

2 Methodology

Brain Tumor Segmentation. We focus on the segmentation of glioblastoma, which are the most common primary malignant brain tumors [1]. A glioblastoma can be segmented into four different compartments: necrosis, edema, contrast-enhancing and non-enhancing tumor. We regard segmentation as a classification problem and rely on four different MRI sequences as input data, which are T_1-, T_1 post-contrast- (T_{1c}), T_2- and FLAIR-weighted images. The sequences are rigidly co-registered, skull-stripped, bias-field corrected [9] and their intensities are normalized via histogram matching to a template image. The four intensity values are stacked in a vector $\mathbf{f}_i \in \mathbb{R}^4$, which in turn is part of a vector image I. The set of voxels in I is denoted by V and the total number of voxels by N. A labeling of I is referred to by $X = \{x_i\}_{i \in V}$ with x_i being a scalar value that indicates the tissue compartment, i.e. $x_i \in \{1, \ldots, m\}$. Seven classes ($m = 7$) including three normal (csf, gray matter, white matter) and the previously mentioned four tumor tissues are defined and contained in a set $\mathcal{L}$.

Dense CRF. We interpret the image I as an undirected graph $G = (V, E)$, where V denotes the set of nodes and E the set of edges. Every node in G is associated with a random variable x_i. The pair (I, X) is a CRF, if for any given I the distribution $P(X|I)$ factorizes according to G. The conditional distribution corresponds to a Gibbs distribution $P(X|I) = \frac{1}{Z(I)} \exp(-E(X|I))$, where $Z(I)$ is the partition function. The most probable (MAP) labeling for a given image I can then be estimated by minimizing the energy. The energy is a sum of potentials $\psi_c(x_c)$ defined for each clique c in G. Pairwise CRFs contain unary potentials ψ_u (=data-likelihood) and pairwise potentials ψ_p (=prior). In contrast to grid-structured CRFs, in dense CRFs pairwise potentials are defined between all pairs of voxels, i.e. G is a complete graph. The energy is given by $E(X|I) = \sum_i \psi_u(x_i) + \sum_{i \sim j} \psi_p(x_i, x_j)$. We define the unary potentials to take the form $\psi_u(x_i) = -\log(p(x_i|h_i))$, where h corresponds to the feature vector proposed by Meier et al. [6] and the posterior probability is estimated by a decision forest. In the approach proposed by Krähenbühl et al. [5] the pairwise potentials ψ_p are restricted to correspond to a weighted mixture of Gaussian kernels. We use a combination of an appearance (using intensities $\mathbf{f}_i$) and smoothness kernel (using voxel coordinates $\mathbf{p}_i$):

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \left\{ \exp(-\tfrac{1}{2}(\mathbf{f}_i - \mathbf{f}_j)^T \Lambda_f (\mathbf{f}_i - \mathbf{f}_j)) + \exp(-\tfrac{1}{2}(\mathbf{p}_i - \mathbf{p}_j)^T \Lambda_p (\mathbf{p}_i - \mathbf{p}_j)) \right\}. \quad (1)$$

The label compatibility function $\mu(x_i, x_j)$ is defined to be a symmetric function that takes into account interactions between all possible combinations of label pairings. The precision matrices of the two kernels are Λ_f and Λ_p.

3 Experiments and Results

In order to assess the segmentation performance of the proposed dense CRF model, we employed publicly-available data of the MICCAI BRATS 2013 & 2015

Challenges [7]. First, a comparison of the dense CRF to a grid-structured CRF was performed, for which we used 20 high-grade cases of the original BRATS 2013 data set for training (18 for training, two for validation purposes). We assessed segmentation performance of both models on the BRATS 2013 challenge set (10 patient cases), the BRATS 2013 leaderboard set (21 cases, excluding low-grade glioma), and a clinical data set of a local university hospital (25 cases). In contrast to the BRATS data sets, the clinical data set was acquired with a standardized MR acquisition protocol (all sequences except FLAIR were acquired with isotropic 1 mm resolution). Second, we evaluated the proposed approach on the BRATS 2015 dataset (247 cases for training and 110 for testing). This should facilitate future comparisons with other segmentation techniques. We report performance measures as tuples (median, range=max-min) for the complete tumor (all tumor labels combined), the tumor core (tumor labels except edema) and the enhancing tumor. Statistically significant differences were assessed using a paired Wilcoxon signed rank test.

The baseline model (G-CRF) relied on a grid-structured pairwise CRF with a Potts prior as used in [6]. Inference was performed using Fast-PD [3]. The dense CRF (D-CRF) was implemented as described in section 2. For making a fair comparison to G-CRF, we set $\mu(x_i, x_j) = \mathbb{1}\{x_i \neq x_j\}$ (Potts prior) and set $\Lambda_f = \lambda_f I_4$ and $\Lambda_p = \lambda_p I_3$ with $I_{(\cdot)}$ being the identity matrix and $\lambda_f, \lambda_p \in \mathbb{R}$. Note that both models employed the same unary classifier (features and decision forest proposed in [6]). We used 100 trees, maximum depth of 18 and the number of candidate features per node was $\sqrt{n}$ with $n = 237$ features. Moreover, we did not perform a parameter learning but hand-tuned the parameter values for both CRF models using the data of the validation set. This way, we aimed at minimizing differences in segmentation performance attributed to a specific parameter learning method rather than the different degree of connectivity of each model. The results of the comparison between G-CRF and D-CRF are shown in Table 1. The median and range of the Dice-coefficients estimated for D-CRF for the BRATS 2015 testing set are (0.847,0.824), (0.647,0.933) and (0.7,0.926) for complete tumor, tumor core and enhancing tumor, respectively.

Table 1: Evaluation results. Performance measures are given as (median, range=max-min). Left tuple: Results for BRATS 2013 Challenge set. Middle tuple: Results for Leaderboard data. Right tuple: Results for clinical data. Statistically significant differences are indicated with $^\star/^{\star\star}$ for $\alpha = 0.05/0.01$.

Region	Dice coefficient	PPV	Sensitivity
Complete tumor (G-CRF)	$(0.825, 0.09)/(0.771, 0.454)/(0.839, 0.54)$	$(0.759, 0.192)/(0.751, 0.538)/(0.766, 0.603)$	$(0.943, 0.197)/(0.969, 0.377)/(0.922, 0.531)$
Complete tumor (D-CRF)	$(0.825, 0.092)/(\mathbf{0.789}, 0.487)/(\mathbf{0.856}^{\star\star}, 0.542)$	$(\mathbf{0.775}, 0.168)/(\mathbf{0.758}^{\star}, 0.561)/(\mathbf{0.772}^{\star\star}, 0.592)$	$(\mathbf{0.953}, 0.239)/(\mathbf{0.979}, 0.39)/(\mathbf{0.938}^{\star\star}, 0.521)$
Tumor core (G-CRF)	$(0.748, 0.507)/(0.745, 0.873)/(\mathbf{0.693}, 0.600)$	$(0.802, 0.364)/(\mathbf{0.746}, 0.886)/(\mathbf{0.741}^{\star\star}, 0.682)$	$(\mathbf{0.715}, 0.686)/(\mathbf{0.888}^{\star\star}, 0.969)/(\mathbf{0.713}^{\star\star}, 0.754)$
Tumor core (D-CRF)	$(\mathbf{0.756}, 0.530)/(\mathbf{0.746}^{\star}, 0.902)/(0.643, 0.651)$	$(\mathbf{0.819}, 0.400)/(0.720, 0.887)/(0.719, 0.609)$	$(0.698, 0.664)/(0.855, 0.963)/(0.659, 0.743)$
Enhancing tumor (G-CRF)	$^\dagger(\mathbf{0.706}^{\star}, 0.442)/(\mathbf{0.677}, 0.843)/(0.579, 0.474)$	$(0.669, 0.320)/(\mathbf{0.644}, 0.859)/(0.508, 0.560)$	$(0.823, 0.624)/(0.923, 0.987)/(0.816, 0.498)$
Enhancing tumor (D-CRF)	$^\dagger(0.705, 0.441)/(0.660, 0.860)/(\mathbf{0.631}^{\star\star}, 0.518)$	$(\mathbf{0.673}, 0.330)/(0.626, 0.901)/(\mathbf{0.542}^{\star\star}, 0.625)$	$(\mathbf{0.873}^{\star}, 0.614)/(\mathbf{0.944}^{\star\star}, 0.995)/(\mathbf{0.866}^{\star\star}, 0.449)$

†The mean for D-CRF was 0.694 and 0.681 for G-CRF.

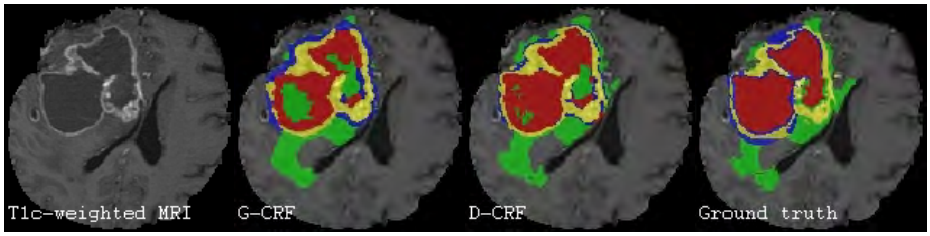


Fig. 1: Exemplary patient case (edema=green, necrosis=red, enhancing tumor=yellow, non-enhancing tumor=blue). From left to right: Contrast-enhanced T_1-weighted Magnetic Resonance sequence. Segmentation result for G-CRF. Segmentation result for D-CRF. Manual ground truth data.

4 Discussion and Future Work

The proposed dense CRF model showed an improved performance for the complete tumor and enhancing tumor segmentation compared to a grid-structured pairwise CRF. This improvement can be attributed to the alleviation of the shrinking bias inherent to pairwise CRFs [5, 2]. The shrinking bias is caused by a combination of a limited neighborhood system and the usage of a Potts prior. The Potts prior encourages smooth boundaries which in grid-structured CRFs results in an oversmoothing of the object boundary. The dense CRF is capable of recovering fine details (cf. Fig. 1 D-CRF, e.g. rim-enhancing tumor) as well as concave object boundaries (e.g. finger-like edematous regions) that otherwise would be smoothed out by a grid-structured pairwise CRF. The local clinical data set was acquired with a higher resolution than the BRATS data, which led to a more significant improvement in performance for D-CRF. However, for segmenting the rather convex-shaped tumor core, oversmoothing of a grid-structured CRF can lead to better results. We plan on further improving our approach by employing a richer prior combined with a parameter learning method [4, 8].

Acknowledgments. This project has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement N°600841.

References

1. von Deimling, A.: Gliomas, Recent Results in Cancer Research, vol. 171. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)
2. Kohli, P., et al: Robust higher order potentials for enforcing label consistency. IJCV 82(3), (2009)
3. Komodakis, N., et al.: Approximate labeling via graph cuts based on linear programming. IEEE TPAMI 29(8), (2007)

4. Krähenbühl, et al.: Parameter Learning and Convergent Inference for Dense Random Fields. ICML (2013)
5. Krähenbühl, P., et al.: Efficient inference in fully connected crfs with gaussian edge potentials. NIPS (2011)
6. Meier, R., et al.: Appearance-and Context-sensitive Features for Brain Tumor Segmentation. MICCAI BRATS Challenge (2014)
7. Menze, B.H., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE TMI, (2014)
8. Orlando, J.I., Blaschko, M.: Learning Fully-Connected CRFs for Blood Vessel Segmentation in Retinal Images. MICCAI (2014)
9. Tustison, N.J., et al.: N4ITK: Improved N3 bias correction. IEEE TMI, (2010)
10. Wang, C., Komodakis, N., Paragios, N.: Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey. CVIU 117(11), (2013)

Multimodal Tumor Segmentation with 3D Volumetric Convolutional Neural Networks

Balaji Pandian[1], Julianne Boyle[1], and Daniel A. Orringer[1]

1. Department of Neurosurgery, University of Michigan Medical School, Ann Arbor, MI, USA, 48109

Abstract. Manual magnetic resonance image (MRI) segmentation is time consuming and has large inter-rater variability. Automatic segmentation of gliomas from MRIs is both highly valuable as well as efficient in the clinical planning of a patients care. With recent computational advances in consumer graphics cards drastically reducing the computational cost of neural networks, we propose a fully automated, novel approach based on a 3D convolutional neural network (CNN) with subvolume training procedures for brain tumor segmentation in multi-modal MRIs (T1, T1c, T2, FLAIR). Our method was tested on a subset of images provided as part of the MICCAI 2016 BRATS challenge. On a training set of 32 patients and a test set of 8 patients, we achieved a preliminary average Dice score of 0.725 for whole tumor segmentation, 0.611 for tumor core segmentation, and 0.572 for active tumor segmentation. We also outline our future steps and experiments, which involve larger training sets, cross-validation, tuning hyperparameters, and optimizing pre and post-processing.

Keywords: Segmentation, CNN, Deep Learning, Glioma, Brain Tumor, Structured Prediction

1 Introduction

Gliomas, a type of brain tumor originating from glial cells, are the most frequent primary brain tumors in adults [2]. Diagnosis, assessment, and treatment planning of gliomas almost always include MRI neuroimaging protocols to evaluate the extent and progression of disease. In current clinical practice, these images are interpreted by either qualitative criteria (for example, the presence of characteristic hyper-intense tissue), or by simple quantitative assessment measures like visible tumor diameter [5]. Improved quantification of the various aspects of a glioma requires accurate segmentation of these portions. Segmentation methods typically look for various regions of gliomas including active tumorous tissue (vascularized or avascular), necrotic tissue, and the peritumoral edematous region. The manual annotation of these boundaries is time-consuming and is prone to misinterpretation and human error [5]. To standardize procedures and ensure high quality care, it would be of high value in neuro-oncology, neurosurgery, and radiotherapy to replace the current manual MRI assessment with automatic, robust, highly accurate, and reproducible segmentation methods.

2 Background and Previous Work

Automated brain MRI tumor segmentation is difficult from a technical standpoint because tumor regions are characterized by intensity changes that are relative to surrounding normal tissue [5]. Furthermore, brain MRIs often display significant intensity variations between subjects, principally when they are acquired with different scanners or at different facilities. Tumor structures also vary greatly across patients in terms of size, shape, extension, and location, impeding the use of strong prior information on shape and location commonly used in the segmentation of many other anatomical structures. This problem of computational brain tumor segmentation has attracted a lot of attention over the past years, resulting in many approaches for the segmentation of tumor structures.

Several groups of the MICCAI 2015 BRATS challenge accomplished high Dice scores using a Convolutional Neural Network (CNN) approach where training is done sequentially on 2D patches in the axial plane. In these methods, segmentation is achieved by predicting the label at the center of a 2D slice. While this has achieved relative success in previous years [1], 3D subvolume convolution has not been attempted due to the significantly increased computational demands. In the past year, recent advances in graphics cards (GPU) and general programming on GPUs have allowed us to extend this 2D patch-based convolutional network into a 3D approach using subvolumes.

We present a fully-automated segmentation method with the intent of accurately segmenting brain tumor regions. We use a 3D CNN to perform the brain tumor segmentation task on the dataset of brain tumor MRI scans provided by the MICCAI 2016 BRATS challenge[1]. The steps of our workflow include: (i) pre-processing, (ii) multimodal 3D CNN feature extraction and 5-category classification (healthy/normal, necrosis, edema, non-enhancing, and enhancing tumor), and (iii) post-processing.

3 Materials

The MRI scans used in this study include data from the BRATS 2012 and BRATS 2013 challenges as well as data from the NIH Cancer Imaging Archive (TCIA) that were prepared as part of the MICCAI BRATS 2014 and BRATS 2015 challenges for a total of 274 cases of patients with gliomas [5]. Due to computational demands and limited time, we used 32 cases for our initial training. The data of each patient is comprised of preoperative multimodal MRIs that were scanned by 4 different modalities including T1, T1 contrast enhanced, T2, and T2 FLAIR MRI. Each scan is a contiguous 3D volume that includes 155 2D slices of size 240x240 pixels. The volumes of the various modalities have already been skull-stripped, co-registered to the same anatomical template, and interpolated to $1mm^3$ voxel resolution. Ground truth segmentations for the enhancing part of the tumor, the tumor core (described by the union of necrotic, non-enhancing and enhancing parts of the tumor), and the whole tumor (which

[1] http://www.braintumorsegmentation.org/

is the union of the tumor core and the peritumoral edematous region) for the training set were also provided.

4 Methods

4.1 Pre-processing

The images from the BRATS 2016 dataset were minimally pre-processed. The N4ITK bias correction [6] was applied to T1 and T1C modalities. Additionally, the data was normalized by shifting and scaling the images so that the voxels in each scan have a zero mean and unit variance.

4.2 Classification

The neural network was designed as a 3D CNN with the following layers:

Table 1. The layout of the 3D CNN

Layer	Type	Parameters	Activation	Input Size	Output Size
1	3D Convolution	32 kernels - size 3x3x3	ReLU	4x9x9x9	32x7x7x7
2	3D Convolution	32 kernels - size 3x3x3	ReLU	32x7x7x7	32x5x5x5
3	Dropout	Probability = 0.1	N/A	32x5x5x5	32x5x5x5
4	3D Convolution	64 kernels - size 3x3x3	ReLU	32x5x5x5	64x3x3x3
5	3D Convolution	64 kernels - size 3x3x3	ReLU	64x3x3x3	64x1x1x1
6	Dropout	Probability = 0.1	N/A	64x1x1x1	64x1x1x1
7	Dense	256 neurons	ReLU	64	256
8	Dense	5 neurons	Softmax	256	5

Loss is calculated using categorical cross-entropy with ADAM as an optimizer [3]. All software was written in Python using Numpy, Keras[2], and Theano[3]. Neural network training and testing was run on a machine with an Intel 4-core i7 CPU, 32GB RAM, and a NVIDIA GTX 1080 GPU.

For each patient in the training set, all possible sub-volumes of size 9x9x9 voxels were enumerated from each of the MRI modalities and stored along with the ground-truth label at the centermost voxel in the sub-volume. An equal number of sub-volumes was selected from each of the 5 classification categories and fed into the CNN for training.

The test set was decomposed into subvolumes using a similar methodology. The main difference between the training and test set is that all 9x9x9 voxel sub-volumes were classified instead of using a predetermined subset of subvolumes.

[2] http://keras.io/
[3] http://deeplearning.net/software/theano/

4.3 Post-processing

For post-processing, we employ the commonly used closing operation. The morphological closing on an image is composed of a dilation operation followed by an erosion operation. Closing can remove small dark spots (pepper noise) and connect small bright cracks. A post-processing method of connected component removal was also implemented to smooth the classifications. The connected components in the volumes are found and removed if they are smaller than the specified size.

5 Results

We randomly chose (without replacement) 32 sets of MRIs as our training set and 8 sets of MRIs for our test set. Using the Dice metric, we achieved an average score of 0.725 for whole tumor segmentation, 0.611 for tumor core segmentation, and 0.572 for active tumor segmentation.

6 Discussion and Conclusion

We present a novel 3D CNN for automatic glioma segmentation that uses overlapping subvolumes instead of 2D patches. The final results show promising preliminary results given the small amount of training data used and the modest amount of pre and post-processing. Future work includes building a classifier from a larger training dataset and tuning the CNN hyperparameters to achieve better classification performance. We believe that these steps will allow us to achieve significantly better Dice scores while maintaining a computational load attainable by consumer personal computers.

References

1. Dvorak, Pavel, and Bjoern Menze: Structured prediction with convolutional neural networks for multimodal brain tumor segmentation. Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge (2015): 13-24.
2. Holland, Eric C.: Progenitor cells and glioma formation. Current opinion in neurology 14.6 (2001): 683-688.
3. Kingma, Diederik, and Jimmy Ba.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
4. Louis, David N., et al.: The 2007 WHO classification of tumours of the central nervous system. Acta neuropathologica 114.2 (2007): 97-109.
5. Menze, Bjoern H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging 34.10 (2015): 1993-2024.
6. Tustison, Nicholas J., et al.: N4ITK: improved N3 bias correction. IEEE transactions on medical imaging 29.6 (2010): 1310-1320.
7. Van Meir, Erwin G., et al.: Exciting new advances in neurooncology: the avenue to a cure for malignant glioma. CA: a cancer journal for clinicians 60.3 (2010): 166-193.

Improving segment boundary classification for Brain Tumor Segmentation and longitudinal disease progression

Ramandeep Randhawa, Ankit Modi, Parag Jain, and Prashant Warier

{ramandeep.randhawa}@marshall.usc.edu
{ankit.modi,prashant.warier}@fractalanalytics.com
{paragjain78}@gmail.com

Abstract. Tracking the progression of brain tumors is a challenging task, due to the slow growth rate and the combination of different tumor components, such as cysts, enhancing patterns, edema and necrosis. In this paper, we propose a DNN based architecture that does automatic segmentation of brain tumor, and focuses on improving accuracy at the edges of these different classes. We show that enhancing the loss function to give more weights to the edge pixels, significantly improves the neural network's accuracy at classifying the boundaries.

1 Introduction

Accurate quantification of gross tumor volumes of brain tumor is an important factor in the assessment of therapy response in patients - it is also important to quantify the volume of the different tumor components, e.g., cysts, enhancing patterns, edema and necrotic regions. In particular, identifying the edges of these tumor components and observing their evolution over time is critical to an accurate assessment of disease progression. Multi-modal MRI is often used to detect, monitor and quantify this progression [1].

Most automatic segmentation models use traditional machine learning approaches - features are manually defined and fed to a classifier, and the algorithms focus on learning the best weights for the classifier. Over the past couple of years, deep learning models have enabled automatic learning of features in addition to the weights used for classification.Several methods using CNNs for brain tumor segmentation have already been proposed [2], [3], [4], [5], [6]. Our work builds upon the work by Pereira et al [4] which uses very small 3x3 kernels, and many convolutional layers, applying several non-linear transformations.

To avoid manual classification between HGG and LGG, we train a single net with the combined image set of HGG and LGG. In addition to standard accuracy measures such as DICE score that determine classification accuracy across the entire segment, we focus our efforts to improve performance at the boundaries between segments. To this end, we propose a weighted cross-entropy loss function, where pixels that are at the boundary of different classes are given

more weight and hence the net learns to classify them better. We visually observe that this approach leads to much better classification of the tumor at the boundaries between different regions. Further, it also provides a modest 1-6% increase in overall dice scores.

2 Dataset

The training dataset of BRATS 2015 [1] comprises brain MRI for 274 patients - each MRI scan has $240 \times 240 \times 155$ voxels. Ground truth is provided for each voxel in terms of one of 5 labels - non-tumor, necrosis, edema, non-enhancing tumor and enhancing tumor. All images are already aligned with T1c and are skull stripped. Tumor regions are defined as:

- Complete (including all 4 tumor structures)
- Core (including all structures except *edema*)
- Enhancing (including the *enhancing tumor* structure only)

Accuracy is measured in terms of Dice score, which is defined as:

$$Dice\ Score = \frac{|P_1 \wedge T_1|}{\frac{1}{2}(|P_1| + |T_1|)}$$

P_1: predicted positive, T_1: ground truth positive

3 Network Architecture

The model processes each 2D slice (axial image) where each pixel has 4 modalities - T1, T1c, T2 and FLAIR. It predicts the class of each pixel by processing a 33x33 patch centered around it - the input to the model is therefore 4x33x33. We used the same net proposed by Pereira et. al. [4], whose network architecture is illustrated in Figure 1. The network uses 8 layers, and has about 28 million parameters. We train the network using a combination of HGG and LGG images, and use a 90/10 split for training/ test data. Further, we use rmsprop [7] to update the network weights during training.

We use two-stage training to train the network - this is done to deal with the fact that the data is highly skewed. The first stage is equiprobable where the training data is randomly sampled equally from all classes - this helps the CONV layers *see* all the classes equally. This is followed by a finetuning stage where we fix the CONV layers and fine-tune fully connected layers, sampling training data according to true class distribution.

In order to improve accuracy of prediction around the edges, we modified the cross-entropy loss function to weigh pixels based on their proximity to pixels of other classes. The weighting function we used is as follows:

$$weight, w = \frac{N + differentCount}{N + sameCount}$$

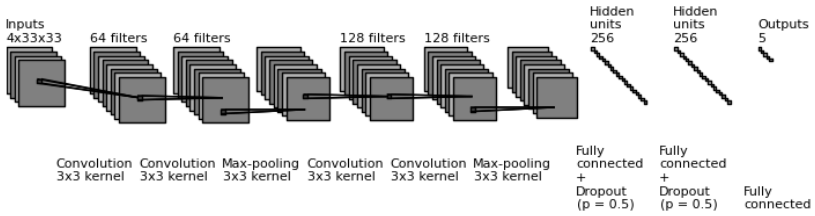


Fig. 1. Network Architecture

where, $differentCount$ is number of pixels that have labels different from the center pixel, $sameCount$ is number of pixels that have the same label and N is a hyper-parameter.
Effectively, the new categorical cross entropy weighted loss used for a minibatch of size j was:

$$Loss = -\sum_{i=1}^{j} x_i w_i \log p_i$$

where, x_i is correct label of data point, w_i is corresponding weight and p_i is corresponding softmax output of the neural network.

4 Results and Discussion

We tried different values of N in the weighting function, and found that values of 1000 - 10,000 work well. At 10,000, the weights range from 0.9 (all pixels are same) to 1.1 (all pixels are different). The results are listed in Table 1 and the visualisations of results are described in Figure 2.

Table 1. Dice scores obtained using the BraTS training dataset; scores on 10% out-of-sample test data

	Complete Tumor (4 tumor classes)	Tumor Core (except edema)	Enhanced Tumor
Baseline Network	0.85	0.72	0.70
Network with Weighted Loss Function	**0.87**	**0.75**	**0.71**

5 Conclusions and Future Work

We intend to extend the architecture with focus on improving performance near edges, enhance learning for difficult predictions, as well as incorporate sagittal

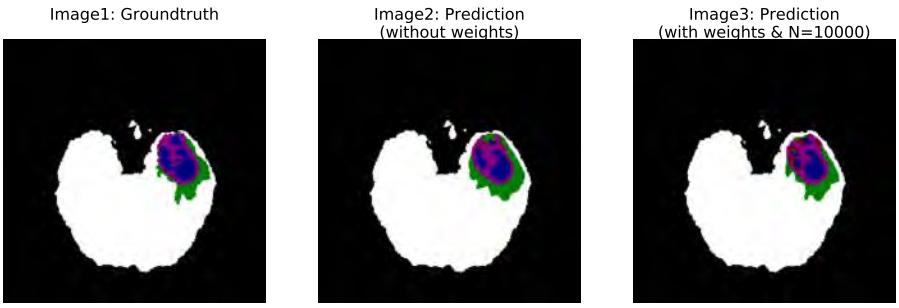


Fig. 2. Groundtruth, prediction without weights and with weights (N=10,000). The images depict 5 classes - the label-color mapping is: necrosis (blue), edema (green), non-enhancing tumor (red), enhancing tumor (pink). In Image 2, one can observe that there are a large number of misclassifications at the edges of the various regions for the baseline setting. Image 3 illustrates that weighted loss function is quite successful at correctly classifying the edge pixels.

and coronal axes. We expect the architecture and results to further improve by the end of the challenge.

References

1. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
2. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, 2016.
3. M. Havaei, F. Dutil, C. Pal, H. Larochelle, and P.-M. Jodoin, "A convolutional neural network approach to brain tumor segmentation," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 195–208, Springer, 2015.
4. S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Deep convolutional neural networks for the segmentation of gliomas in multi-sequence mri," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 131–143, Springer, 2015.
5. G. Urban, M. Bendszus, F. Hamprecht, and J. Kleesiek, "Multi-modal brain tumor segmentation using deep convolutional neural networks. in proc of brats-miccai," 2014.
6. D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, "Segmentation of brain tumor tissues with convolutional neural networks," *Proceedings MICCAI-BRATS*, pp. 36–39, 2014.
7. T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Networks for Machine Learning*, vol. 4, no. 2, 2012.

Glioblastoma Multiforme Segmentation by Variability Characterization of Tumor Boundaries*

Edgar A. Rios Piedra, King-Chung Ho, Ricky K. Taira PhD, Suzie El-Saden MD, Benjamin M. Ellingson PhD, Alex A.T. Bui PhD and William Hsu PhD.

Medical Imaging Informatics Group, Department of Radiological Sciences, University of California, Los Angeles, USA
Department of Radiological Sciences, University of California, Los Angeles, USA
Department of Bioengineering, University of California, Los Angeles, USA

Abstract. Automated medical image analysis plays an important role in diagnoses and treatment assessment, but integration and interpretation across heterogeneous data sources remain significant challenges. In particular, automated estimation of tumor extent in glioblastoma patients has been challenging given the diversity of tumor shapes and appearance characteristics due to differences in magnetic resonance (MR) imaging acquisition parameters, scanner variations and heterogeneity in tumor biology. Here, we present an approach for automated tumor segmentation using multimodal MR images. The algorithm considers the variability arising from the intrinsic tumor heterogeneity and segmentation error to derive the tumor boundary. Using the MICCAI 2015 dataset, a Dice coefficient was obtained of .74 for whole tumor, .54 for tumor core, and 0.54 for active tumor, achieving above average performance in comparison to other approaches evaluated on the BRATS benchmark.

Keywords: Glioblastoma, brain tumor, variability, automatic segmentation.

1 Introduction

Quantitative measurement and assessment of medical images can play an important part in diagnosis of disease, treatment planning, and clinical monitoring. As imaging technology and standards have been rapidly changing, it has become burdensome for clinicians to manually review imaging studies. In addition to increased labor and expense, manual measurements can have a high degree of measurement variability [1] due to the inconsistency and diversity of MRI acquisition parameters (e.g. echo time, repetition time, etc.) and strategies (2D vs. 3D) along with hardware variations (e.g. field strength, gradient performance, etc.) that change the appearance characteristics of the tumor [2]. The increased variability in measurement from multiple imaging sources, combined with the need for faster interpretation, may potentially result in errors associated with treatment decisions or explication of potential therapeutic benefits. We hypothesize the inherent variability in tumor measurements can be leveraged to provide a more accurate assessment of tumor burden. While multiple automated segmentation techniques are being actively developed [3], a method that accounts for the variability in tumor burden estimation remains unexplored.

In this paper, we explore a different perspective towards the identification of tumor boundaries. We developed an iterative algorithm that uses superpixel-based morphological features and prior statistical maps on cerebral tissue distribution to generate a preliminary tumor region. Then, the areas of highest variation inside this preliminary region are obtained to create a Tumor Variability Map (TVM), which represents the image heterogeneity along the tumor boundary (measure of uncertainty).

* Funded by the National Institutes of Health (NIH) under the award number R01CA1575533

2 Methods

The processing pipeline is illustrated on **Pseudo Code 1**. This system is divided into a series of preprocessing strategies followed by the tumor segmentation algorithm. This approach finds an approximate tumor ROI by using a knowledge-based approach. Afterwards, the intensity variation observed on the approximate tumor ROI is analyzed to find the possible tumor boundaries for the TVM. This approach was evaluated using the 2015 Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) dataset [4].

Pseudo Code 1: Brain Tumor segmentation through estimation of variability

for each follow up: d = {d_1,… d_{t-1}, d_t}
    // Data preprocessing
    Read directory and load MRI Sequences → FLAIR, T1, T1+C, T2
    Register and skull strip all volumes;
    Normalize and denoise all volumes;
    Calculate tissue probability masks for white and gray matter → WM, GM
    // Preliminary ROI identification
    for sequences: m = {FLAIR, T2, DeltaMap}
        for orientations x = {axial, coronal, sagittal}
            Obtain brain tissue distributions(WM,GM)
            Cluster input volume(m,x) → Volume superpixels
            Extract image features(m,x) → Histogram, Symmetry, Inhomogeneity
            Find tumor preliminary ROI by using extracted image features
        end
    end
    // Tumor segmentation
    for tumor regions: z = {Edema, Enhancing, Necrosis}
        Get tissue distribution inside ROI (3DtumorROI, WM, GM)
        Evaluate rate of change on tumor region → top regions of intensity variation (t(i))
        Apply thresholds at intensity t(i) inside ROI → Tumor subregion variability map
    end
    Obtain binary tumor masks by obtaining mask consensus on TVM
    Locate unclassifed regions inside tumor area → Non-enhancing tumor
    Warp back all volumes to original scan space
end

2.1 Tumor segmentation

After the images have been preprocessing (including registration, skull stripping, image denoising and intensity normalization), a preliminary tumor ROI is delineated using a knowledge-based approach by obtaining morphological descriptors from each of the input MR sequences that have been partitioned in superpixels using the SLIC algorithm [5]. The resulting ROI is then combined with the context provided by the the approximate distribution of normal cerebral tissues (including gray matter, white matter and cerebrospinal fluid) [6] to obtain the regions with the lowest probability of being normal brain tissues. This process is iterated under different orientations (axial, coronal, sagittal) to increase the accuracy of the initial ROI, hypothesizing that different tumor shapes may be easier for the algorithm to identify if visualized under different perspectives (e.g., a u-shaped tumor might be visualized as two different small structures on the axial view but as a continuous and more defined mass on the coronal view). The initial tumor ROI is then obtained by taking

the union of all regions generated across different perspectives, resulting in a single volumetric ROI.

2.2 Multimodal tumor boundary selection

The next step involves identifying a set of tumor boundaries for the total tumor mass as well as for the tumor sub-regions (i.e., edema, necrosis, enhancing and non enhancing tumor) using the preliminary tumor ROI defined in the previous step. This approach intends to represent the tumor heterogeneous boundary by doing multiple measurements and then combine them into a TVM to quantify uncertainty associated with segmentation boundaries.

The specific tumor boundaries are obtained as follow: A single definition for T2 abnormality was used to define a "T2 abnormal ROI" using the preliminary ROIs found on the FLAIR and T2 contrast images. Regions of edema are extracted by ranking the intensity rate of change on the local ROI histogram, defining as boundaries the locations where the highest total variation across the tumor region was found. A similar process was followed on the T1+C volume or contrast enhanced T1-subtraction-defined (T1+C - T1 volumes) volume to locate the the enhancing and necrotic regions. The TVM is then obtained by aggregating these different approximations of the tumor boundary (**Fig. 2**). A binary representation of the tumor mask is obtained by using the following approach on the TVM:

$$I = \begin{cases} P_{i,j} \geq \dfrac{n}{2} & \therefore \quad I_{i,j} = 1 \\ P_{i,j} < \dfrac{n}{2} & \therefore \quad I_{i,j} = 0 \end{cases}$$

Where I is the output binary image, $P_{i,j}$ is the intensity at pixel location i,j of the TVM P and n is the number of discrete probability levels defined in the variability map. The output is a set of masks that represent the tumor extent and the different sub-regions with the possibility to calculate variability metrics (e.g., agreement ratio, standard deviation, statistical change measurement, and others).

3 Results

This proposed approach was tested on all 220 cases in the BRATS 2015 dataset and evaluated on three components: whole tumor, tumor core (enhancing and necrotic components) and active tumor (enhancing component). The Dice coefficient for total tumor mass of 0.74 (median: 0.77, 1st quartile: 0.66, 3rd quartile: 0.84), 0.54 for the tumor core (median: 0.57, 1st quartile: 0.37, 3rd quartile: 0.75) and 0.54 for the active tumor (median: 0.60, 1st quartile: 0.29, 3rd quartile: 0.76).

4 Discussion

We proposed a multimodal framework for automated, probabilistic brain tumor segmentation by using variability in estimates of the tumor boundary. By exploiting tumor heterogeneity from different imaging sources, this algorithm is able to automatically generate tumor probability maps or alternatively add a measurement of uncertainty to binary tumor segmentations. As the proposed approach iteratively measures the tumor boundaries, it is able to better detect and capture the heterogeneity found on brain tumors (e.g., being able to capture the shape of tumors with eccentric outlines). By explicitly quantifying the error associated with any given segmentation, we believe that this added information is critical to understand and judge the actual tumor extent by a radiologist or neuro-oncologist when interpreting the follow-up imaging data in the clinical setting.

When evaluating the results with other approaches proposed in previous years, our results are comparable to or surpass the mean performance of other algorithms [4] (Reza, Meier, Cordier, Bauer, Festa, Geremia, Buendia, Taylor, Shin). As a method to improve our results, we are also developing a classifier based on Convolutional Neural Networks (CNN) [7] to help in the definition of the preliminary tumor ROI and also to help reduce the number of false positives during the tumor boundary selection. Combining the result of our knowledge-based approach and the result of the CNN (trained to classify whether an individual voxel is part of a brain tumors using an independent dataset) might contribute towards better results on the different tumor contours.

The inclusion of variability calculations into segmentation methodologies can lead to better results and ultimately provide more meaningful data to clinicians as the knowledge of a measurement variation is fundamental to make more objective decisions. Future work includes the evaluation of variability on tumor biomarkers (such as tumor volume, thickness of enhancing margin, necrosis proportion, etc.) and impact on medical decision making.

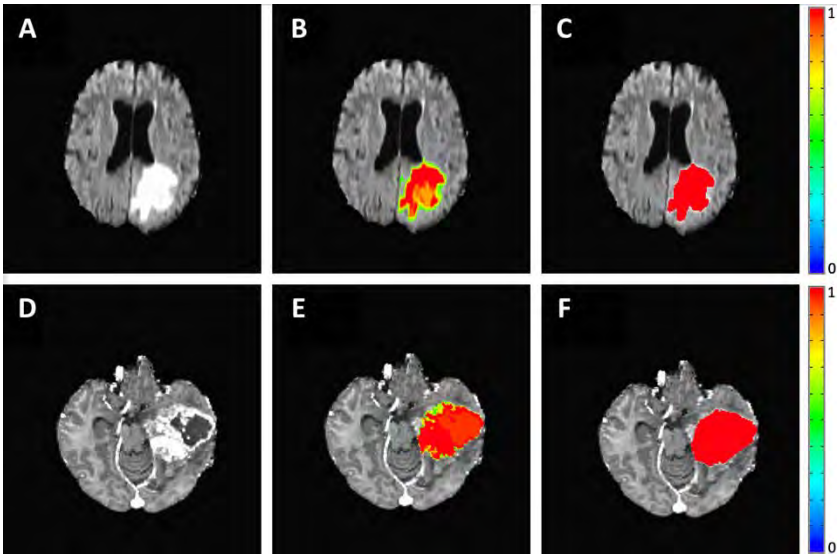


Fig. 2. Examples of brain tumor segmentation results for total tumor mass on two different cases. Showing the original images (A: FLAIR, D: T1+C), the “tumor probability map” (B,E) and binary output for overall abnormality (C,F). The color bar represents the pixels-wise tumor probability.

References

1. Omuro, Antonio, and Lisa M. DeAngelis. "Glioblastoma and other malignant gliomas: a clinical review." *JAMA* 310.17 (2013): 1842-1850.
2. Inda, Maria-del-Mar, Rudy Bonavia, and Joan Seoane. "Glioblastoma multiforme: A look inside its heterogeneous nature." *Cancers* 6.1 (2014): 226-239.
3. Bauer, Stefan, et al. "A survey of MRI-based medical image analysis for brain tumor studies." *Physics in medicine and biology* 58.13 (2013): R97.
4. Menze, Bjoern, Mauricio Reyes, and Koen Van Leemput. "The Multimodal Brain TumorImage Segmentation Benchmark (BRATS)." (2014).
5. Achanta, Radhakrishna, et al. "SLIC superpixels compared to state-of-the-art superpixel methods." *IEEE transactions on pattern analysis and machine intelligence* 34.11 (2012): 2274-2282.
6. Ashburner, John, and Karl J. Friston. "Unified segmentation." *Neuroimage* 26.3 (2005)
7. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

Anatomy-Guided Brain Tumor Segmentation and Classification

Bi Song, Chen-Rui Chou, Albert Huang, and Ming-Chang Liu

US Research Center, Sony Electronics Inc.

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 build classifier for multi-class tumor segmentation by sampling the entire MRI volume data, 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 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.

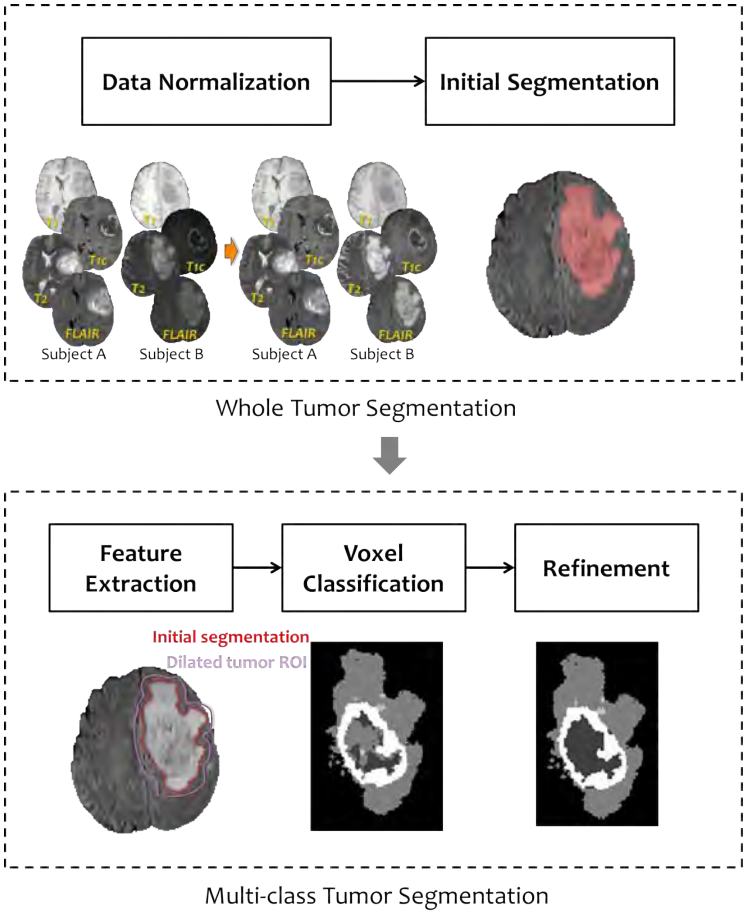


Fig. 1. The proposed two-stage brain tumor segmentation workflow.

2 Method

2.1 Data Normalization

There are large intensity variations across brain tumor MRI data and intensity ranges are not standardized. To normalize the intensity of imaging data, we propose an anatomy-structure-based method based on the assumption that the same structures, such as white matter (WM), grey matter (GM) or cerebrospinal fluid (CSF), should have similar intensity value across different data sets of same modality (T1, T1c, T2, Flair). We apply fuzzy C-means algorithm (FCM) [1, 3] to classify the input data into WM, GM and CSF before normalizing the intensity values of these WM, GM, and CSF classes for each modality.

2.2 Initial Segmentation

Flair and T2 provide better boundary contrast of the whole tumor. By assuming tumor rarely happens completely symmetrically, symmetric differences of Flair and T2 are calculated for locating the initial seeds of tumors. Then the GrowCut algorithm [5] is employed on the linear combination of Flair and T2, $\alpha \cdot I_{Flair} + (1 - \alpha) \cdot I_{T2}$, for segmenting the whole tumor.

2.3 Feature Extraction

The initial segmentation results are dilated to provide aROI for further multi-class segmentation. The features are extracted from the ROI. Our features include voxel-wise and context-wise features.

- *Appearance*: Voxel's intensity value.
- *Texture*: Variance of T2, which represents local inhomogeneity.
- *Context*: Multiscale local mean intensity to catch neighborhood information.

2.4 Voxel Classification

A random forest classifier [2] is used for multi-class segmentation with 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.

2.5 Refinement

Pixel misclassification might error occurred 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.

4

3 Experimental Results

To evaluate the performance of our method, we show results on the BraTS 2015 training data set [4], which contains 220 high-grade and 54 low-grade glioma cases. The data volumes are already skull-stripped and registered intra-patient. Employed accuracy measures are Dice's coefficient and Hausdorff distance. Using a leave-one-out evaluation scheme, we show the obtained results in Table 1 on a total of 274 cases. Note that we do not take high-grade or low-grade as a prior knowledge during training and testing.

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 (%)	85.2	70.0	73.4
Hausdorff Distance (mm)	10.4	11.6	9.1

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. Bezdek, J.: A convergence theorem for the fuzzy isodata clustering algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2(1), 1–8 (January 1980)
2. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
3. Dunn, J.: A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. Journal of Cybernetics 3(3), 32–57 (January 1973)
4. Menze, B.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Image 34(10), 1993–2024 (2015)
5. Vezhnevets, V., Konouchine, V.: GrowCut - Interactive Multi-Label N-D Image Segmentation By Cellular Automata. In: Proc. Graphicon. pp. 150–156 (2005)

3D Convolutional Networks for Brain Tumor Segmentation

Adrià Casamitjana, Santi Puch, Asier Aduriz, Elisa Sayrol, and Verónica Vilaplana*

Signal Theory and Communications Department, Universitat Politècnica de Catalunya. BarcelonaTech, Spain
{adria.casamitjana,elisa.sayrol,veronica.vilaplana}@upc.edu

Abstract. This paper presents our work on applying 3D Convolutional Networks for brain tumor segmentation for the BRATS challenge. We are currently experimenting with different 3D fully convolutional architectures. We present preliminary results using these architectures and outline our future steps and experiments, which involve hyperparameter optimization, comparison of the models' performance and implementation of a post-processing stage to eliminate false positive predictions.

1 Introduction

The problem of automatic brain tumor segmentation has attracted considerable attention during the past years due to its high clinical relevance and its challenging nature [1]. Methods are usually categorized in two broad groups: generative models, which rely on prior knowledge about the appearance and distribution of different tissue types and discriminative models, which directly learn the relationship between image features and segmentation labels.

Within the second group, in the last two years there has been an increasing use of deep learning methods (and specifically convolutional neural networks CNN) to tackle the problem, motivated by the state of the art performance of deep learning models in several computer vision tasks. As opposed to classical discriminative models based on hand-crafted features, deep learning models learn a hierarchy of increasingly complex features directly from data, by applying several layers of trainable filters and optional pooling operations. Most of these methods do not completely exploit the available volumetric information but use two-dimensional CNNs, processing 2D slices independently or using three orthogonal 2D patches to incorporate contextual information (see [1, 3, and references therein]). A fully 3D approach is proposed in [2], consisting on a 3D CNN that produces soft segmentation maps, followed by a fully connected 3D CRF that imposes generalization constraints and obtains the final labels.

* This work has been partially supported by the project BIGGRAPH-TEC2013-43935-R, financed by the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund (ERDF)

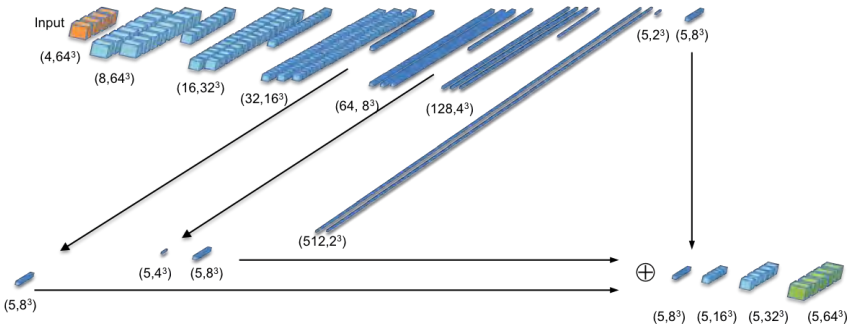


Fig. 1. 3DNet_1

In this work we explore three different 3D CNN architectures to solve the brain tumor segmentation problem. We report preliminary results using the BRATS 2015 Training dataset.

2 Methods

We propose two fully convolutional 3D CNN architectures inspired in two well known 2D models used for generic image segmentation. We also train a third model which is a variant of the two-pathway DeepMedic net proposed in [2].

The first model, 3DNet_1, is a 3D fully convolutional network based on the VGG architecture [8], with skip connections that combine coarse, high layer information with fine, low layer information. The configuration of the net is illustrated in Figure 1. Given the characteristic large amount of parameters of 3D networks, a reduction in the number and dimensions of the filters with respect to its 2D analog was necessary in order to ensure that the model could be trained with the available resources.

The second model, 3DNet_2, is the 3D version of the network proposed in [4]. It is based on the architecture presented in [5], where on top of a VGG net (a contracting path) there is a multilayer deconvolution network (an expansive path). There are connections between corresponding layers in the contracting and expanding paths. The model is illustrated in Figure 2.

The third architecture, 3DNet_3, is a modification of DeepMedic network [2] and is illustrated in Figure 3. The aim of using two paths is gathering both low and high resolution features from the input segment. In a different approach than that of literature, where it is usual to employ different input sizes for each path, we feed segments of equal dimensions to our network. This way, we get coarser features by using greater receptive fields in one path (by means of max-pooling layers) and finer features by using smaller receptive fields (combination of convolutional layers).

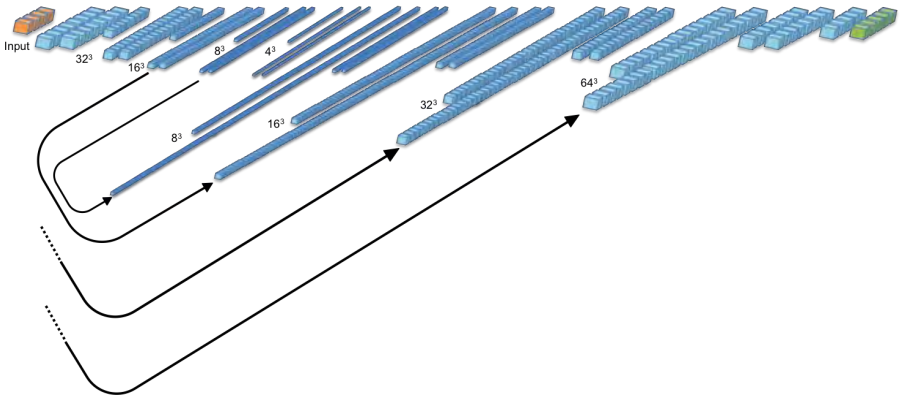


Fig. 2. 3DNet_2

3 Implementation details

Preprocessing: We normalize the data within each input volume by subtracting the volume's mean and dividing by the volume's standard deviation.

Dataset sampling and class imbalance: Due to memory constraints we do not scan the whole volume in one forward pass but divide it into multiple segments of size 32^3. We adopt the training scheme proposed in [2]: training batches are formed by extracting segments from the training images with 50% probability of being centered on a background or foreground voxel. This scheme has the advantage of automatically mitigating the class imbalance problem. In order to further alleviate this problem we weight the cross-entropy loss function taking into account the class distributions as proposed in [7].

Training: All models use ReLu activations and batch normalization. They are trained using the Adam optimizer, with elastic net regularization and using He initialization for the weights. We use a small batch size of 10 due to memory constraints.

4 Results

Preliminary results obtained with the BRATS 2015 Training dataset are presented in Table 1, using 60% of the data for training and the remaining 40% for test. Dice scores are promising for Whole and Core classes, but poor for the Active class, despite the strategies used for dealing with the class imbalance.

5 Conclusions and future work

In this work we explore the use of 3D CNNs for brain tumor segmentation. The three models are fully connected, being capable of dense-inference, that is making predictions for whole volumes in one pass. In addition, the use of

	Accuracy	Dice score			Recall				
		Whole	Core	Active	1-Nec	2-Edm	3-NEnh	4-Enh	0-Else
3DNet_1	0.90	0.81	0.61	0.30	0.48	0.71	0.32	0.75	0.97
3DNet_2	0.95	0.89	0.76	0.37	0.33	0.78	0.45	0.77	0.98
3DNet_3	0.92	0.82	0.70	0.36	0.59	0.69	0.36	0.79	0.98

Table 1. Results for BRATS 2015 Training dataset

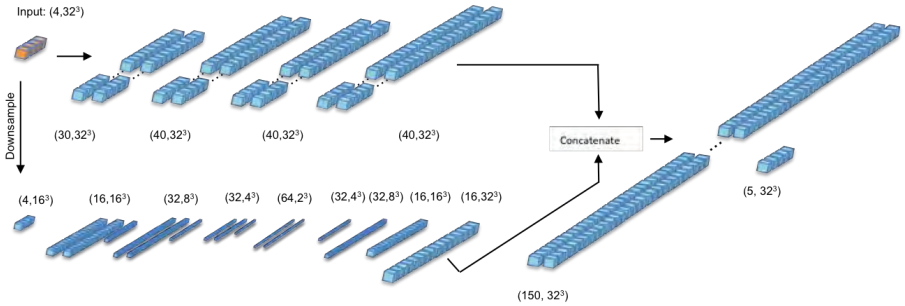


Fig. 3. 3DNet_3

trainable upsampling layers increases the effective batch size without an increase in memory or computational cost.

Future work for the final submission will include: implementation of more elaborate strategies to tackle the class imbalance problem, hyperparameter optimization in order to increase the performance of the models, analysis and comparison of the three architectures and implementation of a post-processing stage to eliminate false positives.

References

1. Menze, B.H., Jakab, A., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). Medical Imaging, IEEE Trans. on 34(10), 1993–2024 (2015)
2. Kamnitsas, K., Ledig, et al.: Efficient Multi-Scale 3D CNN with fully connected CRF for Accurate Brain Lesion Segmentation, arXiv:1603.05959v2 (2016)
3. Havaei, M., Davy, A., et al.: Brain Tumor Segmentation with Deep Neural Networks. arXiv:1505.03540v3 (2016)
4. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1 (2015)
5. Noh, H., Hong, S., Han B.: Learning Deconvolution Network for Semantic Segmentation. ICCV, Santiago, Chile (2015)
6. Long, J., Shelharmer, E., Darrel, T.: Fully Convolutional Networks for Semantic Segmentation. CVPR, Boston, USA (2015)
7. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv:1511.00561 (2015)
8. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556v6 (2015)

Segmentation of Gliomas in Pre-Operative and Post-Operative Multimodal Magnetic Resonance Imaging Volumes Based on a Hybrid Generative-Discriminative Framework

Ke Zeng, Spyridon Bakas, Aristeidis Sotiras, Hamed Akbari, Martin Rozycki, Saima Rathore, Sarthak Pati, and Christos Davatzikos

Section of Biomedical Image Analysis, Center for Biomedical Image Computing and Analytics, Perelman School of Medicine, University of Pennsylvania, USA.
{Ke.Zeng,S.Bakas,Aristeidis.Sotiras,Hamed.Akbari,Martin.Rozycki,
Saima.Rathore,Sarthak.Pati,Christos.Davatzikos}@uphs.upenn.edu

Abstract. We present an approach for segmenting both low- and high-grade gliomas in multimodal magnetic resonance imaging volumes. The proposed framework is an extension of our previous work [4, 5], with an additional component for segmenting post-operative scans. The proposed approach is based on a hybrid generative-discriminative model. Firstly, a generative model based on a joint segmentation-registration framework is used to segment the brain scans into tumor and healthy tissue labels. Secondly, a gradient boosting classification scheme is used to refine tumor labels based on information from multiple patients. We evaluated our approach in 218 cases during the training phase of the BRAin Tumor Segmentation (BRATS) 2016 challenge and report promising results.

Keywords: Segmentation, Brain Tumor, Glioma, Multimodal MRI, Gradient Boosting, Expectation Maximization, Probabilistic Model, BRATS challenge

1 Introduction

Glioma is a common type of brain tumors that originates in the glial cells that surround and support neurons. There are different types of glioma that vary mainly in the degree of their progression rate and histopathology. Commonly, they are classified into low- and high-grade gliomas (LGGs and HGGs).

Gliomas are typically diagnosed by multimodal magnetic resonance imaging (MRI), where two main pathological regions may be identified, the tumor core and the peri-tumoral edematous region. The tumor core typically consists of enhancing (note that LGGs not alway include this part), non-enhancing and necrotic parts. Edema is the response to infiltrating tumor cells, as well as angiogenic and vascular permeability factors released by the spatially adjacent tumor cells [1].

Accurate quantification of these pathological regions may impact significantly treatment decisions, planning, as well as outcome monitoring. However, this is a highly challenging task due to tumor regions being defined through intensity changes relative to the surrounding normal tissue, and such intensity information being disseminated across various modalities for each region. Manually delineating tumor boundaries is a extremely laborious task that is prone to human error and observer bias [6]. Computer-assisted segmentation has the potential to reduce expert burden and raters' variability. Towards this end, we present an interactive segmentation method that aims to accurately delineate gliomas, and eventually allow for their quantification.

The hereby proposed method is an extension of our previously published method called GLISTRboost [4, 5]. While we have restricted our attention to pre-operative scans when designing GLISTRboost, in this work we also consider longitudinal studies and introduce an approach for segmenting gliomas in post-operative scans. The mass effect in many post-operative scans has been substantially relaxed due to the removal of the tumor bulk. Therefore, the component of GLISTRboost that models tumor mass effect is no longer required for segmenting post-operative scans. However, the task of segmenting residual and/or recurrent tumor is made extremely complex due to the presence of surgically-imposed cavities and non-neoplastic contrast enhancements on the cavity's boundaries. To tackle these complexities, we here develop a framework that specifically targets the segmentation of gliomas in post-operative scans. The newly designed framework follows the generative-discriminative principle of GLISTRboost.

The remainder of this paper is organized as follows: Section 2 details the provided data, while Section 3 presents the proposed segmentation strategy. The experimental validation setting is described in Section 4 along with the obtained results. Finally, Section 5 concludes the paper.

2 Materials

The data used in this study comprise 186 pre- and 88 post-operative multimodal MRI scans of patients with gliomas (54 LGGs, 220 HGGs) that were provided as the training set for the multimodal BRATS 2016 challenge. Specifically, these data are a combination of the pre-operative training set (10 LGGs and 20 HGGs) evaluated in the BRATS 2013 challenge [18], 44 LGG and 112 HGG pre-operative scans provided from the National Institutes of Health (NIH) Cancer Imaging Archive (TCIA) and evaluated in the BRATS 2015 challenge, as well as 88 HGG post-operative NIH-TCIA scans that will be evaluated in BRATS 2016. The latter post-operative scans describe longitudinal observations from various time points for 27 patients. Inclusion of these longitudinal scans will allow for the quantitative characterization of these tumors, based on expert neuro-radiologists and the evaluation of the volumetric segmentations, in 'progressing', 'stable disease', or 'shrinking'.

The data of each patient consists of native and contrast-enhanced (CE) T1-weighted, as well as T2-weighted and T2 Fluid-attenuated inversion recovery

(FLAIR) MRI volumes. Finally, ground truth (GT) segmentations for the training set were also provided. Specifically, the data from BRATS 2013 were manually annotated, whereas data from NIH-TCIA were automatically annotated by fusing the approved by experts results of the segmentation algorithms that ranked high in the BRATS 2012 and 2013 challenges [18]. The GT segmentations comprise the enhancing part of the tumor (ET), the tumor core (TC), which is described by the union of necrotic, non-enhancing and enhancing parts of the tumor, and the whole tumor (WT), which is the union of the TC and the peri-tumoral edematous region.

3 Methods

In this section we describe the extended GLISTRboost segmentation framework that addresses tumor segmentation in both pre-operative and post-operative multimodal volumes.

Firstly, all the provided MRI volumes are skull-stripped, co-registered and smoothed to reduce intensity noise in regions of uniform intensity profile [19]. The intensity histograms of all modalities of all patients are then matched to the corresponding modality of a single reference patient. Pre-operative and post-operative volumes are treated differently after the common pre-processing.

A pre-operative scan is segmented by applying GLISTRboost [4, 5], our previously published method that ranked 1st in the BRATS 2015 challenge. GLISTRboost is a hybrid generative-discriminative segmentation framework, in which a generative approach is used to segment the brain scans into tumor, as well as healthy tissue labels. This generative approach is based on a joint segmentation-registration (JSR) scheme [9, 11, 12, 14, 16], whilst incorporating a glioma growth model to deal with the tumor mass effect [10, 13]. A gradient boosting classification (discriminative) scheme is then used to refine tumor labels based on information learned from multiple patients [7, 8]. Lastly, a probabilistic Bayesian strategy, inspired by [2, 3], is employed to further refine and finalize the tumor segmentation based on within-patient brain-specific intensity statistics. Interested readers are referred to [5] for more detailed information.

Post-operative scans are segmented in a slightly different manner than pre-operative scans. Primarily, this is due to the mass effect, which is expected to be substantially relaxed in post-operative scans after the tumor bulk has been resected. Additionally, surgically-imposed cavities and non-neoplastic contrast enhancement on cavity boundaries, none of which is present in pre-operative volumes, substantially complicate the task of segmenting residual and/or recurrent tumors. The generative part of GLISTRboost is adjusted accordingly to address these differences. We firstly remove the glioma growth model from the JSR scheme, with the intention to address the absence of severe mass effect. Taking out this computationally intensive growth model [13] also reduces the time required for the JSR scheme to converge from hours to minutes. Moreover, to address the presence of cavities and enhancement on their boundaries, two additional classes are introduced to the Gaussian mixture model within the JSR

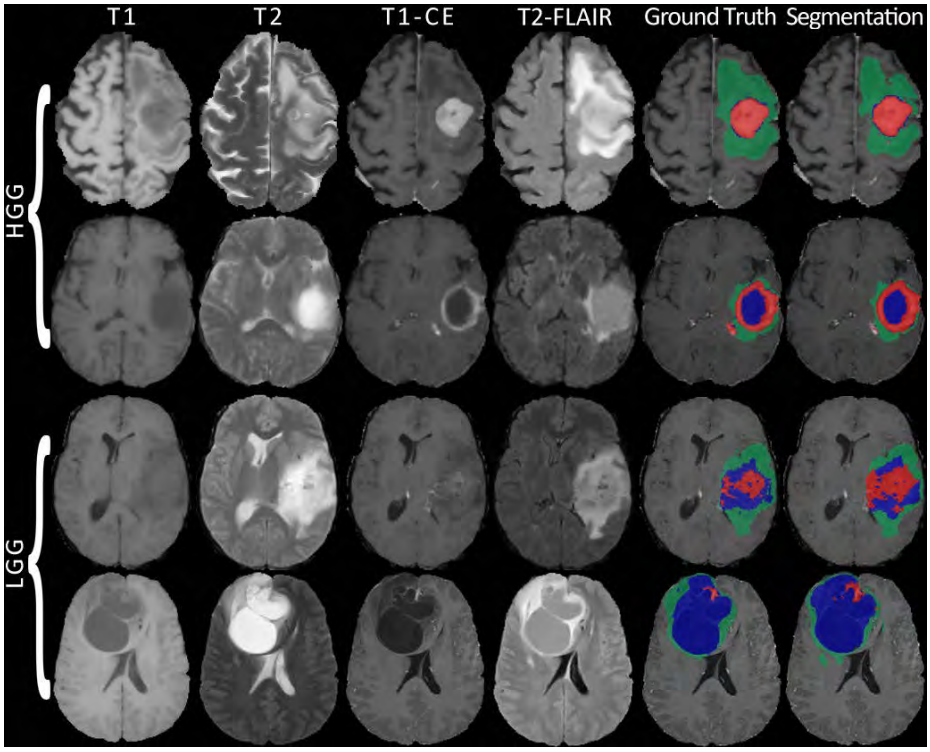
scheme, namely non-enhancing cavity and enhancing cavity. These modifications were also previously considered as a subcomponent of a registration framework that aligns pre-operative to post-operative brain scans of glioma patients [15, 17]. The discriminative step, namely the gradient boosting classification scheme has remained the same as in GLISTRboost. The set of features used for the classifier include two additional features, posterior probability of a voxel in the image being cavity and the geodesic distance of every voxel to the center of the cavity. It should be noted that separate models have been learned for pre-operative and post-operative scans, using only pre-operative and post-operative scans as training samples, respectively.

4 Experiments and Results

In order to assess the segmentation performance of our method in the provided training data, we evaluated the overlap between the tumor labels proposed by our method and the GT in three regions, *i.e.*, WT, TC and ET, as suggested in [18]. Fig. 1.a showcases examples of the available multimodal pre-operative scans, along with their corresponding GT and their segmentation labels as proposed by our method, for four patients (two HGGs and two LGGs). We note the highest overlap in the region of edema and some disagreement between the suggested segmentations and the GT of the enhancing and non-enhancing parts of the tumor. Fig. 1.b illustrates representative post-operative examples of two patients with resection cavities and two with recurrent tumors. We note in the cavity examples, that the cavity itself and its enhancing boundary have almost identical intensity characteristics with the non-enhancing and enhancing parts of the tumor, respectively. As a result, these cavity regions were partially misclassified as these parts of the tumor by the proposed method.

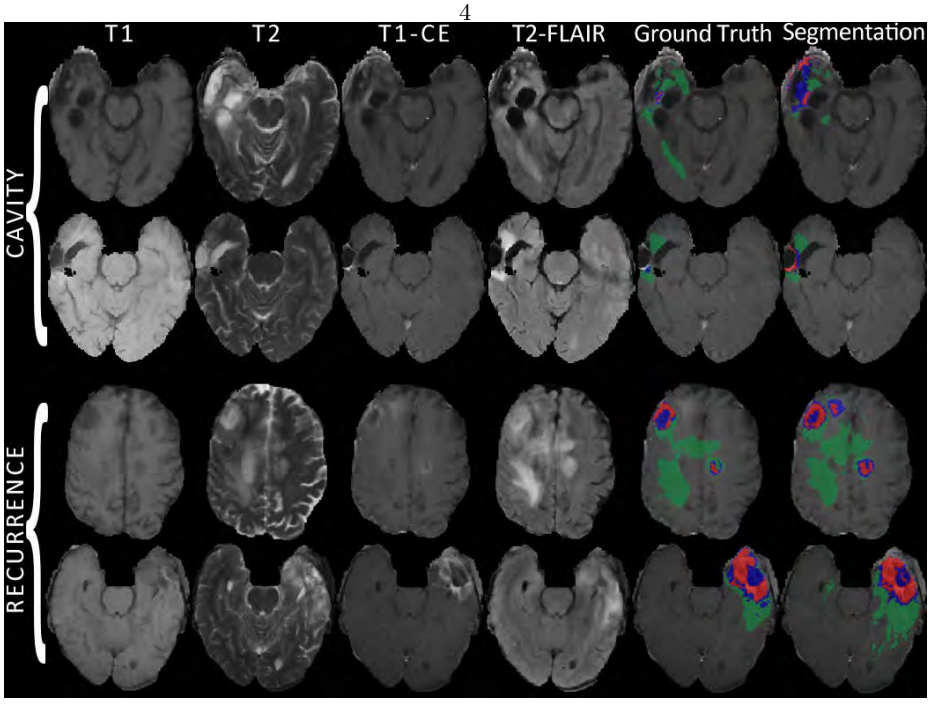
To further appraise the performance of the hereby proposed method, we quantitatively validated the per-voxel overlap between the aforementioned regions (WT, TC, ET) in the training set using the DICE coefficient. Fig. 2 summarizes the distributions of the cross-validated DICE coefficient scores for each tissue label (WT, TC and ET) and across patients of the training set, whilst using only the generative step, as well as the complete method. The results are presented for four different groupings of the data, comprising all pre-operative scans, the pre-operative HGGs, the pre-operative LGGs and the post-operative HGGs. We note a clear improvement after application of the discriminative step in both the mean and median DICE coefficient values for all tissue labels in and data groupings. Furthermore, it is observed that the segmentation results for the ET label varies significantly between LGGs and HGGs, with the former showing lower and less consistent results. This seems to be the effect of training our learning model using both classes simultaneously, when LGGs typically show a different pathophysiological phenotype marked by the lack of an ET. Nevertheless, the segmentation of the WT label in the LGGs is comparable to that of the HGGs. One may also notice that the segmentation quality for post-operative HGG volumes is inferior to and less consistent than that for pre-operative HGG

5



a. Pre-operative multimodal scans.

4



b. Post-operative multimodal scans.

Fig. 1. Representative example segmentation results for four pre-operative and four post-operative multimodal scans. Green, red and blue masks denote the edema, the enhancing tumor and the union of the necrotic and non-enhancing parts of the tumor, respectively.

volumes. Such phenomenon are consistent with the visual inspections in Fig. 1, again highlighting the challenging nature of segmenting post-operative scans. It should be mentioned that even though 88 post-operative HGG scans were included in the training data, we have only used 32 of them for validation purposes due to label misclassifications in their GT. Therefore, we believe that the difference in the number of scans used also affects the segmentation accuracy, and we expect to obtain improved performance by increasing the number of post-operative samples.

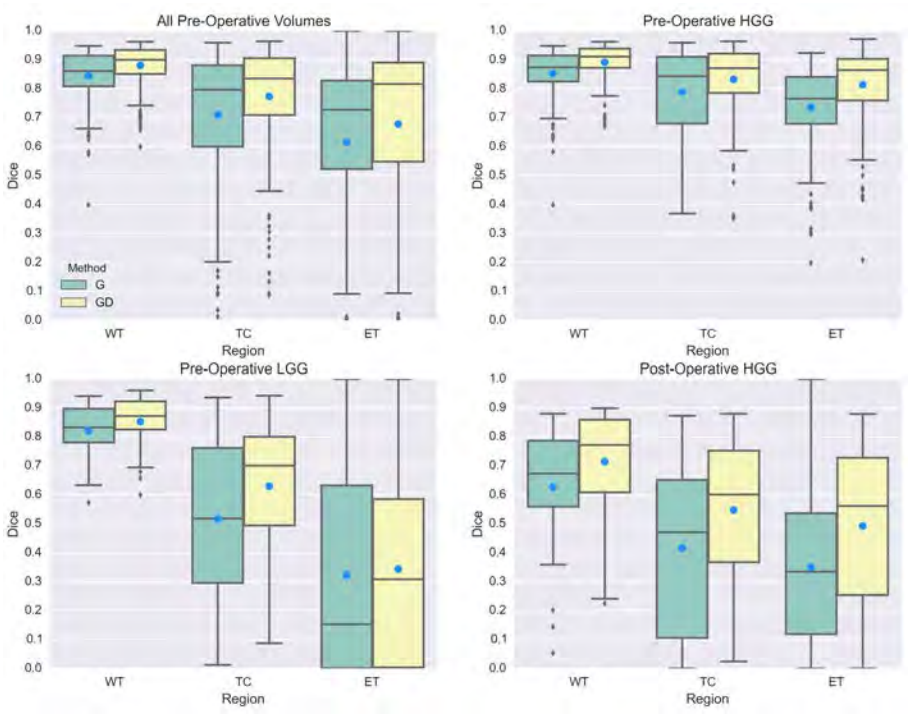


Fig. 2. Distributions of the DICE coefficient across patients for each step of the proposed method, each tissue label and different groupings of data. The different steps are described by the boxes colored in green and yellow for the generative only step (G) and the complete generative-discriminative method (GD), respetively. The dot and the line inside each box denote the mean and median values, respectively.

5 Conclusion

We presented an approach based on a hybrid generative-discriminative framework towards providing a reliable and highly accurate segmentation of gliomas,

in both pre- and post-operative multimodal MRI volumes. The proposed approach is built upon brain segmentation results provided by a generative joint segmentation-registration scheme. Segmentation labels obtained from this generative step are subsequently refined by a discriminative approach, taking into account population-wide tumor label appearance statistics that are learned by employing a gradient boosting classifier.

The proposed approach was validated on the BRATS 2016 training dataset, which is a mixture of pre-operative HGG and LGG scans, as well as post-operative HGG scans. The proposed method ranked first in the BRATS 2015 challenge for producing segmentation labels for gliomas in pre-operative scans. Furthermore, it performs best in all scans for segmenting the WT, whilst for the TC and ET labels it performed best on pre-operative HGGs. Segmentations of the post-operative HGGs were slightly inferior, due to their more challenging nature and fewer available training samples. Overall, the proposed method produced high-quality segmentations on the full dataset and the gradient boosting classifier has substantially improved the segmentation labels in all scenarios.

References

[1] Akbari, H., Macyszyn, L., Da, X., Wolf, R.L., Bilello, M., Verma, R., O'Rourke, D.M., Davatzikos, C.: Pattern analysis of dynamic susceptibility contrast-enhanced MR imaging demonstrates peritumoral tissue heterogeneity. Radiology 273(2), 502–510 (2014)

[2] Bakas, S., Chatzimichail, K., Hunter, G., Labbe, B., Sidhu, P.S., Makris, D.: Fast semi-automatic segmentation of focal liver lesions in contrast-enhanced ultrasound, based on a probabilistic model. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization pp. 1–10 (2015)

[3] Bakas, S., Labbe, B., Hunter, G.J.A., Sidhu, P., Chatzimichail, K., Makris, D.: Fast Segmentation of Focal Liver Lesions in Contrast-Enhanced Ultrasound Data. In Proceedings of the 18th Annual Conference on Medical Image Understanding and Analysis (MIUA) pp. 73–78 (2014)

[4] Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davazikos, C.: Segmentation of gliomas in multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework. In Proceedings of the Multimodal Brain Tumor Image Segmentation Challenge (BRATS) 2015, pages = 5–12, year = 2015,

[5] Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davatzikos, C.: GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation, pp. 144–155. Springer International Publishing, Cham (2016)

[6] Deeley, M.A., Chen, A., Datteri, R., Noble, J.H., Cmelak, A.J., Donnelly, E.F., Malcolm, A.W., Moretti, L., Jaboin, J., Niermann, K., Yang, E.S., Yu, D.S., Yei, F., Koyama, T., Ding, G.X., Dawant, B.M.: Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. Physics in Medicine and Biology 56(14), 4557–4577 (2011)

[7] Friedman, J.H.: Greedy function approximation: A gradient boosting machine. Annals of Statistics 29(5), 1189–1232 (2001)
[8] Friedman, J.H.: Stochastic gradient boosting. Computational Statistics & Data Analysis 38(4), 367–378 (2002)
[9] Gooya, A., Biros, G., Davatzikos, C.: Deformable registration of Glioma images using EM algorithm and diffusion reaction modeling. IEEE Transactions on Medical Imaging 30(2), 375–390 (2011)
[10] Gooya, A., Pohl, K., Billelo, M., Biros, G., Davatzikos, C.: Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI), pages = 532–540, year = 2011,
[11] Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint segmentation and deformable registration of brain scans guided by a tumor growth model. Medical Image Computing and Computer-Assisted Interventions 14(2), 532–540 (2011)
[12] Gooya, A., Pohl, K.M., Bilello, M., Cirillo, L., Biros, G., Melhem, E.R., Davatzikos, C.: GLISTR: Glioma Image Segmentation and Registration. IEEE Transactions on Medical Imaging 31(10), 1941–1954 (2012)
[13] Hogea, C., Davatzikos, C., Biros, G.: An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects. Journal of Mathematical Biology 56(6), 793–825 (2008)
[14] Kwon, D., Akbari, H., Da, X., Gaonkar, B., Davatzikos, C.: Multimodal Brain Tumor Image Segmentation Using GLISTR. MICCAI Brain Tumor Segmentation (BraTS) Challenge Manuscripts pp. 18–19 (2014)
[15] Kwon, D., Niethammer, M., Akbari, H., Bilello, M., Davatzikos, C., Pohl, K.M.: Portr: Pre-operative and post-recurrence brain tumor registration. IEEE Transactions on Medical Imaging 33(3), 651–667 (March 2014)
[16] Kwon, D., Shinohara, R.T., Akbari, H., Davatzikos, C.: Combining Generative Models for Multifocal Glioma Segmentation and Registration. Medical Image Computing and Computer-Assisted Interventions 17(1), 763–770 (2014)
[17] Kwon, D., Zeng, K., Bilello, M., Davatzikos, C.: Estimating Patient Specific Templates for Pre-Operative and Follow-Up Brain Tumor Registration. Medical Image Computing and Computer-Assisted Interventions (2), 222–229 (2015)
[18] Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, L., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Riklin-Raviv, T., Reza, S., Ryan, M., Schwartz, L., Shin, H.C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging p. 33 (2014)
[19] Smith, S.M., Brady, J.M.: SUSAN - a new approach to low level image processing. International Journal of Computer Vision 23(1), 45–78 (1997)

Brain tumor segmentation using a fully convolutional neural network with conditional random fields

Xiaomei Zhao[1] Yihong Wu[1] Guidong Song[2] Zhenye Li[3] Yong Fan[4] and Yazhuo Zhang[2,3,5,6]

[1] National Laboratory of Pattern Recognition,Institute of Automation, Chinese Academy of Sciences
[2] Beijing Neurosurgical Institute, Capital Medical University
[3] Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University
[4] Department of Radiology, Perelman School of Medicine, University of Pennsylvania
[5] Beijing Institute for Brain Disorders Brain Tumor Center
[6] China National Clinical Research Center for Neurological Diseases

Abstract. Deep learning techniques have been widely adopted for learning task-adaptive features in image segmentation applications, such as brain tumor segmentation. However, most of existing deep learning methods are not able to ensure appearance and spatial consistency of segmentation results. In this study, we propose a novel brain tumor segmentation method by integrating a fully convolutional neural network (FCNN) and Conditional Random Fields (CRF), rather than adopting CRF as a post-processing step of the FCNN. We trained our network in 3 steps based on image patches and slices respectively. Our method has achieved promising performance for segmenting tumors based on BRATS 2013 Challenge dataset. Compared with the method ranked first in the challenge, our method is more computationally efficient. Furthermore, our method could achieve competitive performance with only 3 imaging modalities (Flair, T1c, T2), rather than 4 (Flair, T1, T1c, T2), which could reduce the cost of data acquisition and storage.

1 Introduction

It remains a challenging task to automatically segment brain tumors from medical images [1]. Among existing brain tumor segmentation methods, generative models usually acquire prior information through probabilistic atlas image registration[2,3]. However, the image registration is unreliable when the brain is deformed due to large tumors. On the other hand, discriminative models typically segment brain tumors by classifying voxels based on image features. Their segmentation performance is hinged on the image features and classification models. Since deep learning techniques are capable of learning high level and task-adaptive features from training data, they have been adopted in brain tumor segmentation studies. However, most of the existing deep learning methods do not yield segmentation results with appearance and spatial consistency. To

overcome such a limitation, we propose a novel deep network by integrating a fully convolutional neural network (FCNN) and a CRF[4] to segment brain tumors. Our model is trained in 3 steps and is able to segment brain images slice-by-slice, which is much faster than the segmentation method patch-by-patch [5]. Moreover, our method requires only 3 MR imaging modalities (Flair, T1c, T2), rather than 4 modalities (Flair, T1, T1c T2)[5], which could help reduce the cost of data acquisition and storage.

2 The proposed method

2.1 Pre-processing

We firstly use N4ITK [6] to correct the bias field of each MR image. Then, we normalize image intensity by subtracting the gray-value of the highest frequency and dividing the standard derivation of image intensities.

2.2 Brain tumor segmentation model

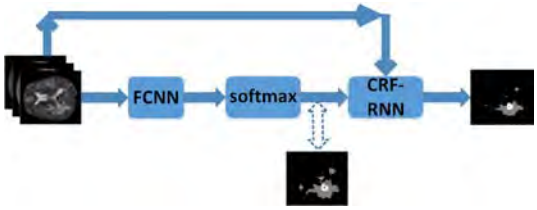


Fig. 1. The structure of our brain tumor segmentation model

Our brain tumor segmentation model consists of 2 parts, a fully convolutional neural network (FCNN) and a Conditional Random Field (CRF). Fig.1 shows the structure of our segmentation model.

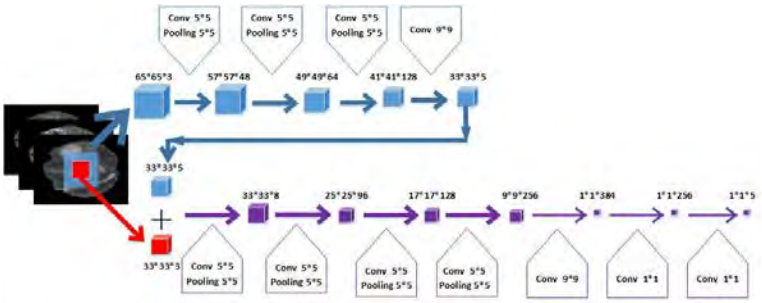


Fig. 2. The structure of our FCNN network

FCNN: Fig.2 shows the FCNN structure. We only use 3 kinds of MR Images in our model, including Flair, T1c and T2. Similar to the method [7], the inputs of our network have two different sizes. However, our network structure and

training process are different from the method proposed in [7]. Our network is a fully convolutional network and the stride of each layer is set to 1.

CRF: CRF could be formulated as a recurrent neural network (RNN), referred to CRF-RNN [4], making it possible to integrate a FCNN and a CRF network as one deep network and train it using a typical back-propagation algorithm.

The combination of FCNN and CRF-RNN: To take full advantage of CRF, the proposed network is trained in 3 steps. Firstly we use image patches to train a FCNN. Then, we use image slices of axial view to train the following CRF-RNN with parameters of the FCNN fixed. Finally, we use image slices to fine-tune the whole network.

2.3 Post-processing

We post-process the segmentation results by removing small 3D-connected regions and correcting some pixels by a simple thresholding method.

3 Experiment

Table 1. Comparison with other methods on BRATS 2013 Challenge dataset

	Dice			PPV			Sensitivity		
	Comp.	Core	Enh.	Comp.	Core	Enh.	Comp.	Core	Enh.
Nick Tustison et al.	0.87	0.78	0.74	0.85	0.74	0.69	**0.89**	**0.88**	**0.83**
Mohammad Havaei et al.[7]	**0.88**	0.79	0.73	0.89	0.79	0.68	0.87	0.79	0.80
Sérgio Pereira et al.[5]	**0.88**	**0.83**	**0.77**	0.88	**0.87**	0.74	**0.89**	0.83	0.81
Mikael Agn et al.[8]	0.87	0.82	0.70	–	–	–	–	–	–
Our method	0.87	0.82	0.76	**0.91**	0.86	**0.77**	0.84	0.81	0.77

Table 2. Comparison with other methods on BRATS 2013 LeaderBoard dataset

	Dice			PPV			Sensitivity		
	Comp.	Core	Enha.	Comp.	Core	Enh.	Comp.	Core	Enh.
Nick Tustison et al.	0.79	0.65	0.53	0.83	0.70	0.51	0.81	0.73	0.66
Mohammad Havaei et al.[7]	0.84	0.71	0.57	**0.88**	0.79	0.54	0.84	0.72	**0.68**
Sérgio Pereira et al.[5]	0.84	**0.72**	**0.62**	0.85	**0.82**	**0.60**	**0.86**	0.76	**0.68**
Mikael Agn et al.[8]	0.83	0.71	0.54	–	–	–	–	–	–
Our method	**0.85**	**0.72**	0.57	**0.88**	0.76	0.59	0.85	**0.77**	0.64

We used BRATS 2013 training dataset as our training data and used BRATS 2013 testing dataset and BRATS 2015 training dataset as our testing data. The experiments were performed on our laboratory's server. The GPU of the server was Tesla K80, and the CPU was Intel E5-2620. As the server was public for everyone in our laboratory, we shared one GPU with other colleagues most time.

Table 3. Comparison with other methods on BRATS 2015 training dataset

Dice	Complete	Core	Enhanced
Mikael Agn et al.[8]	0.77	0.64	0.52
Our method	0.80	0.68	0.65

In generally, it took 2-4 min to segment one subject's imaging data, much faster than the method ranked first on BRATS 2013 Challenge (average running time of 8 min) [5]. Comparison results with other methods are summarized in Tab.1, Tab.2 and Tab.3. Nick Tustison method[1] ranked first in 2013, and Sérgio Pereira method [5] ranked first right now. Since only Mikael Agn et al.[8] used BRATS 2013 training dataset to train their segmentation model and tested their method on BRATS 2015 training dataset, we just compared our segmentation results with Mikael Agn et al.[8] on BRATS 2015 training dataset in Tab.3. These results indicated that our method performed better than Mikael Agn method[8].

Acknowledgements: This work was supported by the National High Technology Research and Development Program of China (2015AA020504) and the National Natural Science Foundation of China under Grant No. 61572499.

References

[1] Bjoern H. Menze et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE transactions on medical imaging, Vol.34, No.10, October 2015, pp. 1993-2024.

[2] Marcel Prastawa, Elizabeth Bullitt, Sean Ho, Guido Gerig. A Brain Tumor Segmentation Framework Based on Outlier Detection, Medical Image Analysis, 8, 2004, pp. 275-283.

[3] Dana Cobzas, Mark Schmidt, Martin Jagersand. 3D Variational Brain Tumor Segmentation using a High Dimensional Feature Set, ICCV 2007,pp. 1-8.

[4] S. Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes et al. Conditional Random Fields as Recurrent Neural Networks, ICCV 2015, pp. 1529-1537.

[5] Sérgio Pereira, Adriano Pinto, Victor Alves and Carlos A. Silva. Brain Tumor Segmentation using Convolutional Neural Networks in MRI Images. In IEEE Transactions on Medical Imaging, Vol. 35, No. 5, 2016, pp. 1240-1251.

[6] Nicholas J. Tustison, Brian B, Avants, Philip A. Cook et al. N4ITK: Improved N3 Bias Correction, IEEE Transactions on Medical Imaging, Vol. 29, No. 6, 2010, pp. 1310-1320.

[7] Mohammad Havaei, Axel Davy, David Warde-Farley et al. Brain Tumor Segmentation with Deep Neural Networks. arXiv: 1505.03540v2 [cs.CV] 5 Oct 2015.

[8] Mikael Agn, Oula Puonti, Ian Law. Brain Tumor Segmentation by a Generative Model with a Prior on Tumor Shape, In Proc. of BRATS-MICCAI 2015.

[1] http://martinos.org/qtim/miccai2013/results.html