Final Report

Karthik Revanuru
IMT2013033

December 11, 2017

Contents

Abstract

Accurate and automated brain tumour segmentation from MRI images is important for performing quantitative analysis of MRI data. There is great potential to utilize quantitative imaging data as prognostic and predictive bio markers for glioma patients. This report details our attempt to apply machine learning techniques to this problem of semantic segmentation posed by BraTS 2017.

1 Introduction

BraTS is a challenge which focuses on the evaluation of state-of-the-art methods for the segmentation of brain tumors in magnetic resonance imaging (MRI) scans. BraTS 2017 utilizes multi-institutional pre-operative MRI scans and focuses on the segmentation of intrinsically heterogeneous (in appearance, shape, and histology) brain tumors, namely gliomas. Furthermore, this year, in order to pinpoint the clinical relevance of this segmentation task, BraTS'17 also focuses on the prediction of patient overall survival, via integrative analyses of radiomic features and machine learning algorithms.

2 Dataset

We are using the dataset published by BraTS[]. Ample multi-institutional routine clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower grade glioma (LGG), with pathologically confirmed diagnosis were provided as the training data. These multimodal scans describe a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes, and were acquired with different clinical protocols and various scanners from multiple (n=19) institutions.

All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the GD-enhancing tumor (ET — label 4), the peritumoral edema (ED — label 2), and the necrotic and non-enhancing tumor (NCR/NET — label 1).

The provided data are distributed after their pre-processing, i.e. co-registered to the same anatomical template, interpolated to the same resolution ($1mm^3$) and skull-stripped. We partitioned the dataset into 70/30 train/test split. We had 285 scans, after the split we had 200 for training and 85 for testing.

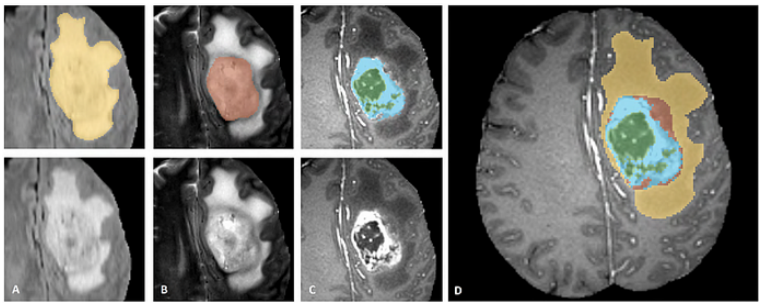


Figure 1: Glioma sub-regions. Shown are image patches with the tumor sub-regions that are annotated in the different modalities (top left) and the final labels for the whole dataset (right). The image patches show from left to right: the whole tumor (yellow) visible in T2-FLAIR (Fig.A), the tumor core (red) visible in T2 (Fig.B), the enhancing tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (Fig. C). The segmentations are combined to generate the final labels of the tumor sub-regions (Fig.D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue)

3 Data Pre-Processing

Though the data was pre-processed i.e co-registered to the same anatomical template, interpolated to the same resolution ($1mm^3$) and skull-stripped, we decided to further pre-process the data after our initial set of experiments. The problem we identified was intensity of pixels is not same across the scans. So we decided to normalize the scans. We have used two methods for normalization.

3.1 Feature Scaling

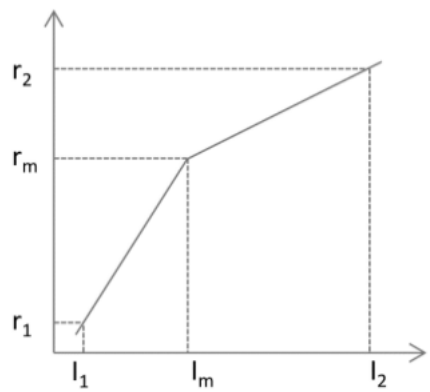
This is normal feature scaling $\frac{X - X_{min}}{X_{max} - X_{min}}$. This would give a value between 0 and 1. We multiply this with 255 to give a value between 0 and 255.

3.2 Landmark Based Normalization

Brain has grey matter and white matter, it means scans will have a binomial distribution. Landmark based normalization was proposed [13] keeping this in view to to make the histogram of MRI images of each modality more similar across different subjects. Landmark is chosen as the intensity value associated with the highest histogram bin of each image ignoring the black background, which typically corresponds to the white matter tissue since it occupies the

largest volume of the brain.

Once the landmark is generated for each MRI image, a linear transform is performed by mapping the landmark intensity to the normalized intensity scale, as shown in the figure below.



X axis represents the intensities in a test image where the landmark has intensity value L_m , and the Y axis denotes the normalized reference intensities. L_1 is taken to be the smallest intensity over the test image and L_2 to be the intensity at the 99.9^{th} percentile within the test image. Beyond that intensities represent outlier values. The landmarks of L_m obtained from the histogram of each image of a subset of images are mapped to the normalized reference intensity scale by linearly mapping the intensities from [L_1,L_2] to [r_1,r_2] in such a way that map L'_m of L_m on [r_1,r_2] can be obtained. The landmark intensity r_m on the reference scale is then determined and fixed as the rounded mean of the L_m's. Once parameters of L_1,L_2 and L_m are obtained, two corresponding slopes of the linear transformation are calculated since parameters of r_1,r_2 and r_m on the reference scale are fixed.

4 Methods

4.1 Segmentation

4.1.1 Initial Approach

In segmentation task we have used random forest[7] as our classifier and we have classified each pixel into one of the four classes (1 for NCR NET, 2 for ED, 4 for ET, and 0 for everything else). We have chosen spatial location of pixel, intensity of each pixel, average of intensities around the pixel over a 5*5*5 and 20*20*20 window and gradient magnitude over a 5*5*5 window. These features

are extracted from various MRI sequences like T1,T2, FLAIR and T1 post contrast images. Idea behind using this is to get both global context and local context at pixel level for classification.

Before giving the entire data to the classifier we have down sampled the data to remove skewness as well as to speed up the training step. We randomly pick pixels for each label so that each class has the same number of data points.

The parameter settings used within the Random Forest algorithm include:

- n estimators = 10 : The number of trees per forest.
- max features = sqrt(n features) : The number of features to consider when looking for the best split.
- min samples split = 2 : The minimum number of samples required to split an internal node.
- min samples leaf = 1 : The minimum number of samples required to be at a leaf node.
- n features = The number of features when fit is performed.

Results obtained using this approach are tabulated in table 1.

4.1.2 Two Step Approach

From the results of our previous approach it's obvious that there is scope for improvement. Basic problem with previous approach was that there is a lot of noise outside the tumour. We tried to remove the noise using different techniques like Gaussian filtering and median filtering but that didn't turn out to be a good idea, as we were loosing out on essential information too. So we have come up with this two step approach.

In the first step taking only Flair images we check each slice of a scan, whether it is a part of tumor or not. We have developed a random forest based classifier for the same, using histogram of intensities as features. After obtaining predictions for all slices we calculate a bounding box around the tumour by smoothing the predictions obtained in the previous step.

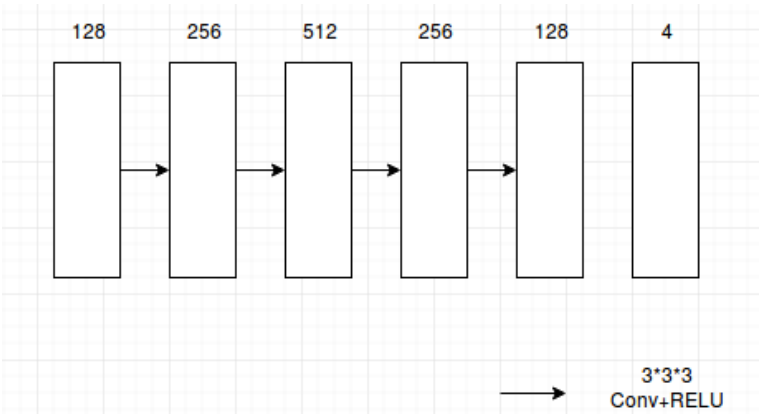
Now we pass only this cut scan to the classifier which is described in the first approach. Intuitively this method should give a boost to accuracy because we are reducing the number of pixels we are classifying by a large factor as tumour is a very small part of a scan and this is possible because we have a rough boundary of tumor. Even after this we see some noise but it's very less compared to what we had before and we remove it using Gaussian filtering. After this prediction we restore the scan into it's original dimensions.

Results obtained using this approach are tabulated in tables 2, 3 and 4.

4.1.3 Karthik-net (K-net)

Convolutional networks are powerful visual models that yield hierarchies of features. Our key insight is to build "fully convolutional" networks that take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. FCN[12] shows that convolutional networks by themselves, trained end-to-end, pixels-to-pixels, exceed the state-of-the-art in semantic segmentation. So we decided to extend upon FCN but as we have only 200 scans we were a little skeptical whether it will actually work as FCN is trained on Imagenet which has huge data. During that time i came across U-Net[11] which presents a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method FCN.

So i decided to experiment with U-Net architecture by implementing it for 3d-segmentation but it has 28 layers and my infrastructure doesn't support for 28 layers so when i was thinking about a way out of this i found max-pool in U-Net architecture to be redundant as i don't need my architecture to be view point invariant because MRI's are guaranteed to be from the same view point. This also prevents the need for Up-sampling layers, so even that loss of information is now not there. I didn't want to go deeper straight away so i started with a simple six layer architecture and named it after me Karthik-net (K-Net) as shown in the figure below. I chose a filter size of 3*3*3 and a stride of 1. And i chose my loss function to be dice score of ground truth and the predicted image. This is again an experiment and i couldn't run this as GPU was down. Code can be found in my repository.



4.2 Survival Prediction

In survival prediction we have used ResNet[9](with soft-max layer removed) to extract features from each of T1,T2, FLAIR and T1 post contrast images. ResNet is a pre-trained network for image recognition which won the 1st place on the ILSVRC 2015 classification task. These extracted features along with age are then passed on to XGBoost[8] for training and prediction.

XGBoost stands for extreme gradient boosting. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance. The features which make this algorithm dominating over others are:

- Sparse Aware implementation with automatic handling of missing data values.
- Block Structure to support the parallelization of tree construction.
- Continued Training so that you can further boost an already fitted model on new data.

Along with XGBoost we have also used Random Forest and Logistic Regression and compared their performance.

5 Experiments and Results

All the experiments were carried out using Scikit learn[10], opencv[6] and Tensor flow[1] on python3.5. Code that has been developed for our experiments can be accessed here[1] and data-set[5][2][3][4] that has been used is provided by the BraTS 2017 organizers.

5.1 Segmentation

For segmentation task currently we have trained on 200 scans and tested on 85 scans we have not included texture as a feature. We have evaluated our methods using dice score and Hausdorff distance. We have also included Sensitivity and Specificity to dice score, allowing us to determine potential over or under segmentation's of the tumor sub-regions. We have also given two sample segmentation's.

[1] `https://github.com/KarthikRevanuru/Brain_Tumor_Segmentation`

5.1.1 Sample Segmentations

Image to the left is an example for a good segmentation. If we observe carefully we can see four different regions clearly. And the image to the right is an example of very bad segmentation it didn't segment the inner regions of the tumour at all and even outside the toumor there are a lot of misclassified labels.

5.1.2 Qunatitative evaluation

Metric	Mean	StdDev	Median
Dice Coefficient	0.40841	0.28784	0.50737
Sensitivity	0.41425	0.33374	0.43963
Specificity	0.97354	0.06178	0.99574
Hausdorff distance	17.09983	10.74695	14.24781

Table 1: Results for our initial approach described in 4.1.1

Metric	Mean	StdDev	Median
Dice Coefficient	0.42027	0.18762	0.47754
Sensitivity	0.43075	0.25949	0.49467
Specificity	0.96334	0.06186	0.9792
Hausdorff distance	13.70209	4.52344	12.56924

Table 2: Results for our two step approach described in 4.1.2, images were not normalized.

5.2 Survival Prediction

Even for survival prediction we have divided the given data into 70 and 30 for training and testing respectively. The results for both classification and regression are tabulated below.

Metric	Mean	StdDev	Median
Dice Coefficient	0.77215	0.11485	0.78361
Sensitivity	0.71957	0.17706	0.72417
Specificity	0.99385	0.01334	0.99886
Hausdorff distance	8.86304	3.91588	8.77478

Table 3: Results for our two step approach described in 4.1.2, images were normalized using feature scaling approach described 3.1

Metric	Mean	StdDev	Median
Dice Coefficient	0.50204	0.26835	0.54769
Sensitivity	0.48952	0.31611	0.50294
Specificity	0.98392	0.01868	0.98959
Hausdorff distance	13.1515	3.74122	12.36932

Table 4: Results for our two step approach described in 4.1.2, images were normalized using landmark based approach described in 3.2.

5.2.1 Regression

Regressor	Mean Avg Error in days
Logistic Regression	253
XGBoost	225
Random Forest	203

5.2.2 Classification

We have used XGBoost and Random Forest for classification.

.	Short	Medium	Long
Short	32	28	16
Medium	8	12	12
Long	20	24	28

The table above is a confusion matrix using XGBoost Classifier.

.	Short	Medium	Long
Short	40	20	16
Medium	8	12	12
Long	20	28	24

The table above is a confusion matrix using Random Forest Classifier.

6 Conclusion and future work

In this paper we have presented our approach for brain tumour segmentation and survival prediction using Random Forests and CNN. Continuing this work we want to study and experiment with different deep learning techniques for segmentation task and also train the K-Net architecture and see how it works. Though we are not satisfied with the results obtained for survival task, we don't want to spend more time on that as it's really difficult even for radiologists or doctors to predicts how many days a patient can survive with decent accuracy. Going further we would also compare our methods with existing baseline methods like Glister and we want to develop a pre-processing pipeline to make our classifier robust to any kind of scans.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. In *Nature Scientific Data*, 2017.

[3] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. In *The Cancer Imaging Archive*, 2017.

[4] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. In *The Cancer Imaging Archive*, 2017.

[5] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber

MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, and Van Leemput K. The multimodal brain tumor image segmentation benchmark (brats). In *IEEE Transactions on Medical Imaging 34(10)*, pages 1993–2024, 2015.

[6] G. Bradski. Open cv. *Dr. Dobb's Journal of Software Tools*, 2000.

[7] Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001.

[8] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[10] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12:2825–2830, November 2011.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[12] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1605.06211, 2016.

[13] Ying Zhuge, Andra V Krauze, Holly Ning, Jason Y Cheng, Barbara C Arora, Kevin Camphausen, and Robert W Miller. Brain tumor segmentation using holistically-nested neural networks in mri images. *Medical Physics*, 2017.