Patient-Specific Semi-supervised Learning for Postoperative Brain Tumor Segmentation

Raphael Meier[1], Stefan Bauer[1,2], Johannes Slotboom[2], Roland Wiest[2], and Mauricio Reyes[1]

[1] Institute for Surgical Technologies and Biomechanics, University of Bern, Switzerland
[2] Inselspital, Bern University Hospital, Switzerland
raphael.meier@istb.unibe.ch

Abstract. In contrast to preoperative brain tumor segmentation, the problem of postoperative brain tumor segmentation has been rarely approached so far. We present a fully-automatic segmentation method using multimodal magnetic resonance image data and patient-specific semi-supervised learning. The idea behind our semi-supervised approach is to effectively fuse information from both pre- and postoperative image data of the same patient to improve segmentation of the postoperative image. We pose image segmentation as a classification problem and solve it by adopting a semi-supervised decision forest. The method is evaluated on a cohort of 10 high-grade glioma patients, with segmentation performance and computation time comparable or superior to a state-of-the-art brain tumor segmentation method. Moreover, our results confirm that the inclusion of preoperative MR images lead to a better performance regarding postoperative brain tumor segmentation.

1 Introduction

Brain tumors are a rather rare but fatal disease. The most common type of primary brain tumors are gliomas, where the Glioblastoma (GBM) is its most aggressive form with a median patient survival that ranges from 12.2 to 15.9 months [1].

The current approach for treatment of glioma patients involves primary tumor surgery (resection) followed by combined radio- and chemotherapy. The imaging modality of choice is Magnetic Resonance Imaging (MRI). Recent clinical studies, such as e.g. [2], use manual, image-based volumetric analysis rather than diameter-based measures for assessing the outcome of tumor surgery. Lately, it has been shown that manual segmentation of postoperative GBM images is being subject to large interobserver variability [3]. Fully-automatic segmentation methods have the potential to resolve this issue.

High-grade gliomas such as GBMs can be subdivided into four different tumor subcompartments: enhancing tumor, non-enhancing tumor, necrosis and edema [5]. We are interested in segmenting residual enhancing tumor since knowledge about its location and volume is of great clinical relevance. In radiation therapy the enhancing tumor is used for defining the gross tumor volume to be

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 714–721, 2014.

targeted. Furthermore, the volume of residual enhancing tumor serves as an inclusion criteria for chemotherapy [4] and correlates with patient survival [2]. The segmentation of postoperative brain tumor images is more challenging than segmenting preoperative images for various reasons such as:

- Hemorrhages (caused by surgery) may appear hyperintense on T_1-weighted MR images which can lead to confounding with enhancing tumor.
- Depending on the amount of blood degradation products contained in the resection cavity, the appearance of the cavity can be confounded with the appearance of necrosis or edema.
- The appearance of the postoperative image is influenced by factors that can not be straightforwardly included in a computational model (such as e.g. the experience and skill of the neurosurgeon). From a statistical learning-based point of view these additional (when compared to preoperative images), external influences can be seen as additional dimensions of our feature space. Informally, one can then say that for having the same predictive performance on postoperative images as for preoperative images, a larger number of training samples is needed (curse of dimensionality).

In this work, we rely on multimodal MR images for discriminating hemorrhages from enhancing tumor. We do not attempt to model external influences on the image appearance. However, we try to minimize the overall complexity by employing a machine learning-based model that is *patient-specific*. In other words, we try to solve the present image segmentation by direct inference (i.e. solving it for one patient at a time) rather than induction (i.e. inferring a 'general' rule).

The problem of segmenting postoperative brain tumor images (compared to preoperative images, e.g. see [6]) has received little attention so far. In 2002, Moonis et al. [7] proposed a method based on fuzzy-connectedness for segmenting postoperative MR images. Their approach is semi-automatic and requires the manual definition of seed points. Kanaly et al. [8] proposed a semi-automatic method based on thresholding of the difference image between T_1-weighted pre- and postcontrast images. Recently, Kwon et al. [9] developed a preoperative and post-recurrence brain tumor registration, which included the segmentation of (postoperative) post-recurrence images using a Bayesian joint registration and segmentation framework. In contrast, we segment the initial postoperative scan where recurrences are usually absent and circumvent the need for time-consuming registration procedures via the use of semi-supervised learning. Semi-supervised learning has been used previously in the context of brain tumor segmentation. Lee et al. [10] proposed a semi-supervised discriminative random field. They also employed a patient-specific model for segmenting enhancing tumor. However, they applied their model only on two-dimensional, preoperative images. Caban et al. [11] proposed a framework of sequential transductive and inductive learning based on conditional mixture Naive Bayes and Support Vector Machines. Their framework was designed for annotating edema in multi-modal, temporal MRI studies of patients with high-grade gliomas.

Our contribution is a fully-automatic method for segmenting the enhancing tumor in postoperative multimodal MR images. We introduce a patient-specific

semi-supervised learning approach that is resilient to the presence of hemorrhages by combining information from pre- and postoperative images without the need of registering them.

2 Methods

We approach the problem of segmenting postoperative multimodal MR images of brain tumor patients from a machine learning-based point of view. Hence, we consider it as a classification problem in which we seek a hypothesis h that maps a voxel in an image to its corresponding tissue class label. Voxels are represented by a feature vector $\mathbf{x} \in \mathbb{R}^n$. The target tissue class label $y \in \{0, 1\}$ is a binary variable, representing enhancing tumor ($y = 1$) and remaining brain tissue ($y = 0$) respectively. Furthermore, for every patient we are given a preoperative multimodal image $\Omega_{pre} = \{\omega_{T1}, \omega_{T1c}, \omega_{T2}, \omega_{FLAIR}\}$ and a postoperative image Ω_{post} also consisting of T_1-weighted, T_1-weighted post-contrast, T_2-weighted and $FLAIR$-weighted MR images. Those four modalities are considered standard in clinical acquisition protocols. We further rely on a training set $\mathcal{S}$, which will be used to infer h (=training). A previously unseen voxel i can then be classified via $h(\mathbf{x}^{(i)}) : \mathbf{x}^{(i)} \rightarrow y^{(i)}$ (=testing).

2.1 Features

Before extracting voxel-wise feature vectors, a multimodal image is preprocessed. This step encompasses noise-reduction, intensity normalization and bias field correction (corresponds to the pipeline proposed in [12]). The features to be extracted can be subdivided into appearance- and context-sensitive features. Appearance-sensitive features are the voxel-wise monomodal intensity values, voxel-wise intensity difference between pre- and post-contrast T_1-weighted images, first-order statistics (extracted over a 26-voxel neighborhood) and gradient magnitude textures (local mean and variance) of the respective modalities. Context-sensitive features are symmetric intensity differences computed between the contralateral hemispheres. The axis of symmetry has been defined as the mid-sagittal plane in an atlas. For increasing the robustness of the symmetric features, we smooth the images with a Gaussian kernel ($\sigma = 3.0$) before extracting them. In the end, we obtain a 45-dimensional feature vector $\mathbf{x}$.

2.2 Patient-Specific Semi-supervised Learning

For solving our classification problem, we make two assumptions: First, we assume that for every preoperative image Ω_{pre} a corresponding label map can be generated. Ideally, such a map consists of labels for the healthy tissues (CSF, GM, WM) and four tumoral subcompartments (enhancing tumor, non-enhancing tumor, necrosis and edema). Second, we assume that the enhancing tumor and its residual appear *sufficiently* similar in the pre- and postoperative images (implying proximity in the feature space).

Regarding the first assumption, considerable improvement on segmenting preoperative brain tumor images has been achieved the last two years (see the MICCAI BRATS challenges 2012 & 2013 [5]). Hence, such a label map can be automatically created by a segmentation algorithm. We think that the second assumption holds for pre- and postoperative images of the same patient. Those two assumptions form together with the requirement that the postoperative image Ω_{post} to be segmented is available during training the basis of our approach.

In supervised learning, we are given a fully-labeled training set $\mathcal{S} = \{(\mathbf{x}^{(i)}, y^{(i)}): i = 1, ..., |\mathcal{S}|\}$ ($|\cdot|$ representing the cardinality of a set), whereas in a semi-supervised setting only a subset ($\mathcal{S}_\ell \subseteq \mathcal{S}$) of the training data is labeled. This setting can now be translated to our situation, where we have for every patient j a fully-labeled preoperative image $\Omega_{pre,j}$ and an unlabeled postoperative image $\Omega_{post,j}$. The main idea is now to train a model both on the labeled preoperative image data as well as on the unlabeled postoperative data of the *same* patient. This way, information from the pre- and postoperative image can be combined through a common feature space, omitting an error-prone and time-consuming registration step. The final aim is that the model estimates for every voxel in the postoperative image the corresponding tissue class label. Since test data (postoperative image) is already available during training the labels can be propagated from the labeled to unlabeled data, which is also known as *transduction*. Consequently, our transductive model tries to solve the classification problem directly for the available data rather than inferring a general rule h.

2.3 Semi-supervised Decision Forest

Decision Forests are an increasingly popular discriminative model mainly used for solving classification and regression problems. A thorough introduction into decision forests in the context of computer vision and medical image analysis can be found in the book of Criminisi et al. [13]. In [13] a semi-supervised variant of the decision forest model has been proposed, yet only applied to two-dimensional toy datasets. We adopt this model for solving our problem of postoperative brain tumor segmentation.

During training of a supervised decision forest, for each decision tree, data is passed down from the root to the leafs. In doing so, data is split such that for every internal (=split) node k of the respective decision tree an objective function is maximized. We consider the information gain IG_k defined as

$$IG_k(\mathcal{S}_k, \theta_k) = H(\mathcal{S}_k) - \sum_{i \in \{L,R\}} \frac{|\mathcal{S}_k^i|}{|\mathcal{S}_k|} H(\mathcal{S}_k^i) \tag{1}$$

where $H(\mathcal{S}_k)$ denotes the entropy, $\mathcal{S}_k^i$ the training data after the split and $\{L, R\}$ index the left and right child node respectively.

In every split node k, we choose the parameters θ_k of a weak learner $h_k(\mathbf{x}, \theta_k)$ such that the information gain is maximized, i.e. $\theta_k^\star = \arg\max_{\theta_k \in \Theta} IG_k(\mathcal{S}_k, \theta_k)$, where Θ denotes the parameter space. In other words, an optimal split is a decision boundary which separates the training data such that the resulting

empirical class distributions of the children show minimal entropy. Optimization is performed by exhaustive search over a randomly selected subspace of Θ. In this work, we consider h_k to be an axis-aligned hyperplane. For handling unlabeled training data the information gain defined in equation (1) is extended by an unsupervised term $IG_{k,u}$ resulting in

$$IG_k(\mathcal{S}_k, \theta_k) = IG_{k,u}(\mathcal{S}_k, \theta_k) + \alpha \cdot IG_{k,s}(\mathcal{S}_{k,\ell}, \theta_k) \tag{2}$$

where $IG_{k,s}$ corresponds to equation (1), $\mathcal{S}_{k,\ell}$ denotes the labeled subset of the training data in node k and $IG_{k,u}$ is defined as

$$IG_{k,u}(\mathcal{S}_k, \theta_k) = \log(\det \Sigma(\mathcal{S}_k)) - \sum_{i \in \{L,R\}} \frac{\left|\mathcal{S}_k^i\right|}{|\mathcal{S}_k|} \log(\det \Sigma(\mathcal{S}_k^i)). \tag{3}$$

The sets $\mathcal{S}_k$, $\mathcal{S}_k^i$ refer to the complete (labeled and unlabeled) training data before and after the split and Σ is a $n \times n$ covariance matrix, respectively. The coefficient α controls the influence of the labeled data. Above formulation corresponds to the one proposed in [13].

A decision forest establishes a partitioning of the feature space. Since we are considering axis-aligned weak learners, we obtain rectangularly shaped partitions. A regular decision forest uses data in such a partition to estimate the respective leaf statistics. In the particular case of semi-supervised decision forests we are dealing with unlabeled data as well as labeled data. This has the consequence that some of the data points or even all data points affiliated with a leaf are unlabeled. However, for estimating leaf statistics all the training data points need a label. Thus, we have to propagate class labels from the labeled subset in a meaningful way. As suggested in [13], this can be realized by finding the closest labeled point in terms of a geodesic distance G. In the present case, G corresponds to the shortest geodesic path along data points, starting and ending at an unlabeled and labeled data point respectively. The local distance function between data points is chosen to be the symmetric Mahalanobis distance [13] (results from the assumption that the leaf density follows a multivariate Gaussian distribution, cf. equation (3)). Since performing label propagation for each data point separately is computationally not feasible, the forest model approximates it by conducting it for the leaf centroids only. Therefore, we compute the geodesic distance between the means associated with the gaussian partitions. Leafs are represented as nodes in a graph. The shortest (discrete) geodesic distance can then be determined by solving the all-pairs shortest path problem for this particular graph, which can be achieved by the Floyd-Warshall algorithm. The closest leaf statistics are then propagated to the respective unlabeled leaf.

After propagation of leaf statistics, the posterior probability $p(y|\mathbf{x}_u)$ for an unlabeled feature vector corresponds to an average over the whole forest: $p\left(y|\mathbf{x}_u\right) = 1/T \sum_{t=1}^{T} p_t\left(y|\mathbf{x}_u\right)$, where $p_t\left(y|\mathbf{x}_u\right)$ is computed based on the empirical class-histogram stored in the leaf of tree t containing vector $\mathbf{x}_u$. The assignment of an unlabeled voxel i to its most probable tissue class $\tilde{y}^{(i)}$ is then performed according to the MAP-rule: $\tilde{y}^{(i)} = \arg\max_y p(y^{(i)}|\mathbf{x}_u^{(i)})$.

3 Results

For evaluating our method, we relied on image data of 10 high-grade glioma patients (images resampled to 1 $[mm]$ isotropic resolution, mean preoperative contrast-enhancing tumor volume in $[ml]$: 13.7 ± 11.2). This encompasses pre- and postoperative multimodal images of four modalities ($T_1, T_{1c}, T_2, FLAIR$). For five patients a complete resection of contrast-enhancing tumor has been performed, whereas for the other five patients a residual tumor volume is present. Before the evaluation, all the images were skullstripped and rigidly registered. The (pre- and postoperative) ground truth was defined by manual expert segmentation.

For comparison, we chose one of the top-ranked segmentation methods (which is based on [12]) of the BRATS Challenge [5]. The method, primarily designed for segmenting preoperative low- and high-grade gliomas, employs a supervised decision forest classifier followed by a spatial regularization using a Conditional Random Field. We think that due to the methodological similarity between our approach and the one chosen for comparison, the influence of semi-supervised learning may be more apparant. Class imbalance is taken into account via undersampling of the majority class ($y = 0$). Our method, which we refer to as *SSDF* (Semi-Supervised Decision Forest), is trained and evaluated on the labeled preoperative and unlabeled postoperative image of one patient at a time. The supervised method, which we simply refer to as *DF*, is once trained on labeled preoperative images only (*DFPRE*), once trained and evaluated (leave-one-out cross-validation) on labeled postoperative images (*DFPOST*) and finally trained and evaluated using both labeled pre- and postoperative images (leave-one-out cross-validation) (*DFPREPOST*).

Our method has been implemented using C++ and the Sherwood library [13]. We fixed the number of trees T to 40 and α to 1.0. The computation time is mainly defined by the number of leafs, i.e. by the depth D of the forest, and the size of the image volume. The number of leafs l correspond to the number of nodes in the all-shortest path problem to be solved for label propagation. Since the Floyd-Warshall algorithm has a time complexity of $O(l^3)$, computation time increases drastically with D. We found that a depth of $D = 8$ resulted in a good trade-off between performance and computation time.

For quantitative evaluation of the segmentation results, we chose to measure sensitivity, specificity, positive predictive value (PPV) and absolute volume error in $[ml]$. We chose not to estimate the Dice coefficient due to the small size of the residual tumor segments, which results in drastic changes of overlap measures. Results are depicted in table 1. Sensitivity and PPV can be computed for the patients with residual tumor volume only. Specificity and absolute volume error for all 10 patients. We further defined a true positive value of zero for a particular method and image to be a 'miss' and counted the total number of missed instances (#MISSED) per method. The average computation time of *SSDF* (training + testing) is about 3.5 minutes, which is less than the average testing time for *DF* of about 5 minutes.

Table 1. Results of quantitative evaluation. Performance measures are described by the tuple (median, range) due to the small sample size.

Method	Sensitivity	Specificity	PPV	Abs. volume error $[ml]$	#MISSED
SSDF	$(0.16, 0.27)$	$(0.99, 0.08)$	$(0.24, 0.93)$	$(0.24, 4.72)$	1
DFPRE	$(0.26, 0.61)$	$(0.96, 0.26)$	$(0.15, 0.93)$	$(2.38, 11.24)$	0
DFPOST	$(0, 0.12)$	$(0.99, 0.01)$	$(0, 0.92)$	$(0.15, 6.49)$	3
DFPREPOST	$(0.19, 0.26)$	$(0.99, 0.06)$	$(0.25, 0.92)$	$(0.48, 5.87)$	1

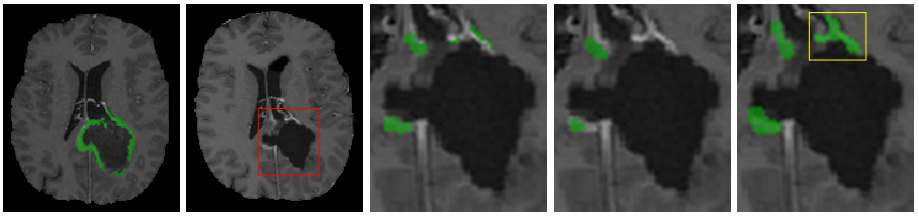

Fig. 1. Exemplary case, T_{1c}-weighted images (green = enhancing tumor), from left to right: Preoperative image, Postoperative image (relevant area is magnified), result for *SSDF*, result for *DFPREPOST*, ground truth. In this specific case, part of the choroid plexus (yellow box) has been infiltrated by the tumor which was correctly detected by our approach.

4 Discussion and Conclusion

When analyzing the segmentation results qualitatively as well as quantitatively, we make several observations. First, *DFPRE* tends to oversegment the residual enhancing tumor. This is reflected in the high sensitivity, low specificity (compared to other approaches), highest absolute volume error and zero misses. The oversegmentation is caused by the inability of the model to properly discriminate between hemorrhage and enhancing tumor. In contrast, *DFPOST* tends to undersegment the residual enhancing tumor and completely misses three cases. The performance of our approach *SSDF* seems superior to *DFPRE* which indicates that in other preoperative images lies no relevant information for segmenting a particular postoperative image. The results of the combined model *DFPREPOST* are comparable to *SSDF* which suggests that information from pre- and postoperative images is truly complementary for segmenting postoperative images. In favor of *SSDF* is the smaller error in volume. An exemplary case supporting our idea of patient-specificity is shown in figure 1. However, due to the small sample size these observations are of preliminary nature. The advantage of *SSDF* over *DFPREPOST* is that it can discriminate residual enhancing tumor from hemorrhages without requiring postoperative ground truth data. The generation of ground truth data for postoperative images is subject to large inter-observer variability [3] and even more time-consuming than for preoperative images.

We presented a fully-automatic method for segmenting residual enhancing tumor in postoperative MR images, which is important to assess surgical outcome and make further treatment decisions. Our approach allows information from pre- and postoperative images to be integrated without the necessity to register them. We think that our initial results provide a basis for further research of this important yet rarely approached segmentation problem.

Acknowledgments. This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No [600841], from the Swiss Cancer League and the Swiss National Science Foundation.

References

1. Zinn, P.O., Colen, R.R.: Imaging genomic mapping in glioblastoma. Neurosurgery 60, 126–130 (2013)
2. Sanai, N., Polley, M.Y., McDermott, M.W., et al.: An extent of resection threshold for newly diagnosed glioblastomas. J. Neurosurg. 115(1), 3–8 (2011)
3. Kubben, P.L., Postma, A.A., Kessels, A.G.H., van Overbeeke, J.J., van Santbrink, H.: Intraobserver and interobserver agreement in volumetric assessment of glioblastoma multiforme resection. Neurosurgery 67(5), 1329–1334 (2010)
4. Newton, H.: Handbook of Brain Tumor Chemotherapy. Academic Press (2006)
5. B. Menze, et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) (submitted, 2014)
6. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of MRI-based medical image analysis for brain tumor studies. Phys. Med. Biol. 58(13), 97–129 (2013)
7. Moonis, G., Liu, J., et al.: Estimation of tumor volume with fuzzy-connectedness segmentation of MR images. AJNR 23(3), 356–363 (2002)
8. Kanaly, C.W., Ding, D., Mehta, A.I., Waller, A.F., Crocker, I., Desjardins, A., Reardon, D.A., Friedman, A.H., et al.: A novel method for volumetric MRI response assessment of enhancing brain tumors. PLoS One 6(1) (2011)
9. Kwon, D., Niethammer, M., Akbari, H., Bilello, M., Davatzikos, C., Pohl, K.: PORTR: Pre-Operative and Post-Recurrence Brain Tumor Registration. IEEE TMI 33(3), 651–667 (2014)
10. Lee, C.H., Wang, S., et al.: Learning to Model Spatial Dependency: Semi-Supervised Discriminative Random Fields. In: Schölkopf, B., Platt, J., Hofmann, T. (eds.) NIPS 2006, vol. 19, pp. 793–800. MIT Press, Cambridge (2006)
11. Caban, J.J., Lee, N., Ebadollahi, S., Laine, A.F., Kender, J.R.: Concept Detection in Longitudinal Brain MR Images Using Multi-Modal Cues. In: IEEE ISBI, pp. 418–421. IEEE Press, New York (2009)
12. Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.P., Reyes, M.: Segmentation of Brain Tumor Images Based on Integrated Hierarchical Classification and Regularization. In: Proceedings of MICCAI-BRATS 2012, pp. 10–13 (2012)
13. Criminisi, A., Shotton, J.: Decision Forests for Computer Vision and Medical Image Analysis. Springer, London (2013)