# Segmenting Glioma in Multi-Modal Images using a Generative-Discriminative Model for Brain Lesion Segmentation

Bjoern H. Menze[1,2], Ezequiel Geremia[2], Nicholas Ayache[2], and Gabor Szekely[1]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland
[2] Asclepios Research Project, INRIA Sophia-Antipolis, France

## 1 Introduction

In this paper, we evaluate a generative-discriminative approach for multi-modal tumor segmentation that builds – in its generative part – on a generative statistical model for tumor appearance in multi-dimensional images [1] by using a "latent" tumor class [2, 3], and – in its discriminative part – on a machine learning approach based on a random forest using long-range features that is capable of learning the local appearance of brain lesions in multi-dimensional images [4, 5]. The approach combines advantageous properties from both types of learning algorithms: First, it extracts tumor related image features in a robust fashion that is invariant to relative intensity changes by relying on a generative model encoding prior knowledge on expected physiology and pathophysiological changes. Second, it transforms image features extracted from the generative model – representing tumor probabilities in the different image channels – to an arbitrary image representation desired by the human interpreter through an efficient classification method that is capable of dealing with high-dimensional input data and that returns the desired class probabilities. In the following, we shortly describe the generative model from [1], and input features and additional regularization methods used similar to our earlier discriminative model from [4].

## 2 Generative Tumor Model

We use a generative modeling approach, in which we first build an explicit statistical model of image formation and subsequently use this model to derive a fully automatic segmentation algorithm. We follow closely our description of the method from [1]. The structure of the generative probabilistic model provides strong priors on the expected image patterns. For segmenting magnetic resonance (MR) images, it has the advantage that model parameters describing the observed image intensities serve as nuisance parameters. This makes it robust against tissue specific changes of the image intensity, and the algorithm does not depend on intensity calibration methods – often required for learning approaches that use image intensities as input – that may be prone to errors in the presence of lesions that vary in size and image intensity. Moreover, generative image appearance model can be combined with other parametric models

used, for example, for registration [6] or bias field correction [7], and even more complex image modification that can be "regressed out" in the same way.

*Tumor appearance model* We model the *normal state* of the healthy brain using a spatially varying probabilistic prior $\boldsymbol{\pi}_k$, a standard population atlas, for each of the $K = 3$ tissue classes that are visible from the given images (gray matter, white matter, and cerebrospinal fluid). The normal state $k_i$ is shared among all $C$ channels at voxel $i$. We model the *tumor state* using a spatially varying "latent" probabilistic atlas $\boldsymbol{\alpha}$, similar to [2]. At each voxel $i$, this atlas provides a scalar parameter $\alpha_i$ that defines the probability of observing a tumor transition at that voxel. Parameter $\alpha_i$ is unknown and is estimated as part of the segmentation process. We further assume that *image observations* $y_i^c$ are generated by Gaussian intensity distributions for each of the $K$ tissue classes and the $C$ channels, with mean $\mu_k^c$ and variance $v_k^c$, respectively. If the image in channel $c$ shows a transition from normal tissue to tumor in voxel $i$ (i.e., if tissue state $s_i^c = T$), the normal observations are replaced by intensities from another set of channel-specific Gaussian distributions with mean $\mu_T^c$ and variance $v_T^c$, representing the tumor class.

*Biological constraints on the estimated parameters* We seek Maximum Likelihood estimates of the model parameters $\{\boldsymbol{\theta}, \boldsymbol{\alpha}\}$ by estimating the tissue state vector $\boldsymbol{s}_i$ of every voxel $i$ that indicates the type of tissue visible in the different image modalities. The vector has $\boldsymbol{s}_i^c = T$ in all channels that show tumor, and has value $s_i^c = k_i$ in all channels that appear normal. With $K = 3$ tissues types and $C = 4$ channels (for the given data), the cardinality of the state vector is $|\boldsymbol{s}| = K * 2^C = 3 * 2^4 = 48$. However, plausibility constraints on the expected tumor appearance in the different channels apply, for example ruling out tumor-induced intensity changes in T1gad unless the same location also shows tumor-induced changes in both T2 and FLAIR, and only gray and white matter being able to show tumor transitions, the number of possible tissue states reduces to $|\boldsymbol{s}| = 7$. We estimate the most likely state vector $\boldsymbol{s}_i$ in a standard expectation maximization procedure, similar to the "EM segmention" algorithm, with iterative updates of the parameters $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\alpha}}\}$ and the posterior probabilities $p(\boldsymbol{s}_i|k_i, \boldsymbol{y}_i; \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\alpha}})$. Updates can be performed using intuitive closed-form expressions: the latent tumor prior $\tilde{\alpha}_i$ is an average of the corresponding posterior estimated, and the intensity parameters $\tilde{\mu}^c$ and $\tilde{v}^c$ are updated with the weighted statistics of the data for the healthy tissues and for the tumor class. During the iteration we enforced that tumor voxels are hyper- or hypo-intense with respect to the current average gray value of the white matter tissue (hypo-intense for T1, hyper-intens for T1gad, T2, FLAIR) similar to [8]. Also we encourage smoothness of the tumor labels by extending the latent atlas $\alpha$ to include a Markov Random Field (MRF) prior, relaxing the MRF to a mean-field approximation with an efficient approximate algorithm. Different from [1], we now use channel-specific regularization parameters $\beta$ that are all in the range of .3 to 1. Typically convergence is reached after 10-15 updates.

*Channel-specific tumor and tissue probabilities* Once we have an estimate of the model parameters $\{\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\}$, we can evaluate the probability that tumor is visible in channel $c$ of voxel $i$ by summing over all the configurations $\boldsymbol{S}_i$ for which $S_i^c = T$:

$$p(s_i^c = T|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) = \sum_{\boldsymbol{t}_i} \delta(s_i^c, T)\ p(\boldsymbol{t}_i|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}), \tag{1}$$

where $\delta$ is the Kroneker delta that is 1 for $s_i^c = T$ and 0 otherwise. The generative model returns $C$ tumor appearance map $p(s_i^c = T|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}})$, one for each channel of the input volume. It also returns the probability maps for the $K$ healthy tissues $p(k_i|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}})$, with global estimates for each voxel $i$ that are valid for all $C$ images.

## 3 Discriminative Lesion Model

The present generative model returns probability maps for the healthy tissues, and probability maps for the presences of characteristic hypo- or hyper-intens changes in each of the image volumes. While this provides highly specific information about different pathophysiological processes induced by the tumor, the analysis of the multimodal image sequence may still require to highlight specific structures of the lesion – such as edema, the location of the active or necrotic core of the tumor, "hot spots" of modified angiogenesis or metabolism – that cannot directly be associated with any of these basic parameter maps returned. As a consequence, we propose to use the probabilistic output of the generative model, together with few structural features that are derived from the same probabilistic maps, as input to a classifier modeling the posterior of the desired pixel classes. In this we follow the approach proposed by [4] that prove useful for identifying white matter lesion in multiple input volumes. The building blocks of this discriminative approach are the input features, the parametrization of the random forest classifier used, and the final post-processing routines.

*Image features* As input feature describing the image in voxel $i$ we use the probabilities $p(k_i)$ for the $K = 3$ tissue classes ($\boldsymbol{x}_i^k$). We also use the tumor probability $p(s_i^c = T)$ for each channel $C = 4$ ($\boldsymbol{x}_i^c$), and the $C = 4$ image intensities after calibrating them with a global factor that has been estimated from gray and white matter tissue ($\boldsymbol{x}_i^{im}$). From these data we derive two types of features: the "long range features" that calculate differences of local image intensities for all three types of input features ($\boldsymbol{x}_i^k$, $\boldsymbol{x}_i^c$,$\boldsymbol{x}_i^{im}$), and a distance feature that calculates the geodesic distance of each voxel $i$ to characteristic tumor areas.

The first type of features calculate the difference between the image intensity, or scalar of any other map, at voxel $j$ that is located at $\boldsymbol{v}$ and the image intensity at another voxel $k$ that is located at $\boldsymbol{v} + \boldsymbol{w}$ (with $\boldsymbol{v}$ here being 3D spatial coordinates). For every voxel $j$ in our volume we calculate these differences $\boldsymbol{x}_j^{diff} = \boldsymbol{x}_j - \boldsymbol{x}_k$ for 20 different directions $\boldsymbol{w}$ around $\boldsymbol{v}$ with spatial offsets

between 3mm to 3cm. To reduce noise the subtracted value at $\boldsymbol{v}+\boldsymbol{w}$ is extracted after smoothing the image intensities locally around voxel $k$ (using a Gaussian kernel with 3mm standard deviation).

The second type of features calculates the geodesic distance between the location $\boldsymbol{v}$ of voxel $j$ to specific image feature that are of particular interest in the analysis. The path is constrained to areas that are most likely gray matter, white matter or tumor as predicted by the generative model. More specifically, we use the distance of $\boldsymbol{x}_j^{\delta tissue}$ of voxel $j$ to the boundary of the the brain tissue (the interface of white and gray matter with CSF), and the distance $\boldsymbol{x}_j^{\delta edema}$ to the boundary of the T2 lesion representing the approximate location of the edema. The second distance $\boldsymbol{x}_j^{\delta edema}$ is calculated independently for voxels outside $\boldsymbol{x}_j^{\delta edema+}$ and inside $\boldsymbol{x}_j^{\delta edema-}$ the edema. In total we have 269 image features $\boldsymbol{x}$ for each voxel when concatenating the vectors of $\boldsymbol{x}^k$, $\boldsymbol{x}^c$, $\boldsymbol{x}^{im}$, $\boldsymbol{x}^{diff}$, and $\boldsymbol{x}^{\delta}$.

*Classifier and spatial regularization* We use the image features $\boldsymbol{x}$ defined above to model the probabilities $p(L;\boldsymbol{x})$ of class labels $L$ in the BRATS data set, and the labels of the $K$ normal tissue. For the normal classes (that are not available from the manual annotation of the challenge data set) we infer the maximum a posterior estimates of the generative model and use them as label during training. We choose random forests as our discriminative model as it uses labeled samples as input and returns class probabilities. Random forests learn many decision trees from bootstrapped samples of the training data, and at each split in the tree they only evaluate a random subspaces to find the best split. The split that separates samples of one class best against the others (with respect to Gini impurity) is chosen. Trees are grown until all nodes contain sample of a single class only. In prediction new samples are pushed down all trees of the ensemble and assigned, for each tree, to the class of the terminal node they end up in. Votes are averaged over all trees of the ensemble. The resulting normalized votes approximate the posterior for training samples that are independent of each other [9]. To minimize correlation in the training data, and also to speed up training, we draw no more 2000 samples from each of the $\approx 10^6$ voxels in each of the 25 data set. We train an ensemble with 300 randomized decision trees, and choose a subspace dimensionality of 10. We use the random forest implementation from Breiman and Cutler. To improve segmentation, we use a Markov Random Field (MRF) imposing a smoothness constraint on the class labels. We optimize the function imposing costs when assigning different labels in a 6 neighbourhood on the cross-validated predictions on the training data.

## 4 Experiments

We evaluate our model on the BRATS challenge data set of 25 patients with glioma. The data set comprises $T_1$, $T_2$, FLAIR-, and post-Gadolinium $T_1$ MR images, all images are skull stripped and co-registered. We segment the volume

into the three healthy and an outlier class using a freely available implementation of the EM segmentation with bias correction [7, 8]. Outliers are defined as being more than three standard deviations away from the centroid of any of the three normal tissue classes. We apply the generative model to the bias field corrected volumes and initialize intensity parameters with values estimated in the initial segmentation; we also use the the bias field and intensity corrected images as input for the discriminative model. More details about these data is given in another submission to the BRATS challenge that focuses on evaluating the generative model [10].

Exemplary segmentations that are returned from the present approach are shown in Figure 1 and quantitative results from a leave-one-out cross-validation are shown in Table 1. Note that the definition of "core" labels differs between ground truth (where it also includes the T1 hypo-intense center of the tumor) and the algorithm tested (where it is only the T1gad hyper-intense area of the tumor) which results in misleading evaluation scores for the "core" class in low-grade cases.

**Acknowledgements.** This work was supported by NCCR Co-Me of the Swiss National Science Foundation, and INRIA CompuTumor.

## References

1. Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.A., Ayache, N., Golland, P.: A generative model for brain tumor segmentation in multi-modal images. In: Proc MICCAI. (2010) 151–159
2. Riklin-Raviv, T., Menze, B.H., Van Leemput, K., Stieltjes, B., Weber, M.A., Ayache, N., Wells, W.M., Golland, P.: Joint segmentation via patient-specific latent anatomy model. In: Proc MICCAI-PMMIA (Workshop on Probabilistic Models for Medical Image Analysis). (2009) 244–255
3. Riklin-Raviv, T., Van Leemput, K., Menze, B.H., Wells, 3rd, W.M., Golland, P.: Segmentation of image ensembles via latent atlases. Med Image Anal **14** (2010) 654–665
4. Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel MR images. In: Proc MICCAI. (2010)
5. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. Neuroimage **57** (2011) 378–90
6. Pohl, K.M., Warfield, S.K., Kikinis, R., Grimson, W.E.L., Wells, W.M.: Coupling statistical segmentation and -pca- shape modeling. In: Medical Image Computing and Computer-Assisted Intervention. Volume 3216/2004 of Lecture Notes in Computer Science., Rennes / St-Malo, France, Springer, Heidelberg (2004) 151 – 159
7. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based bias field correction of MR images of the brain. IEEE T Med Imaging **18** (1999) 885–896

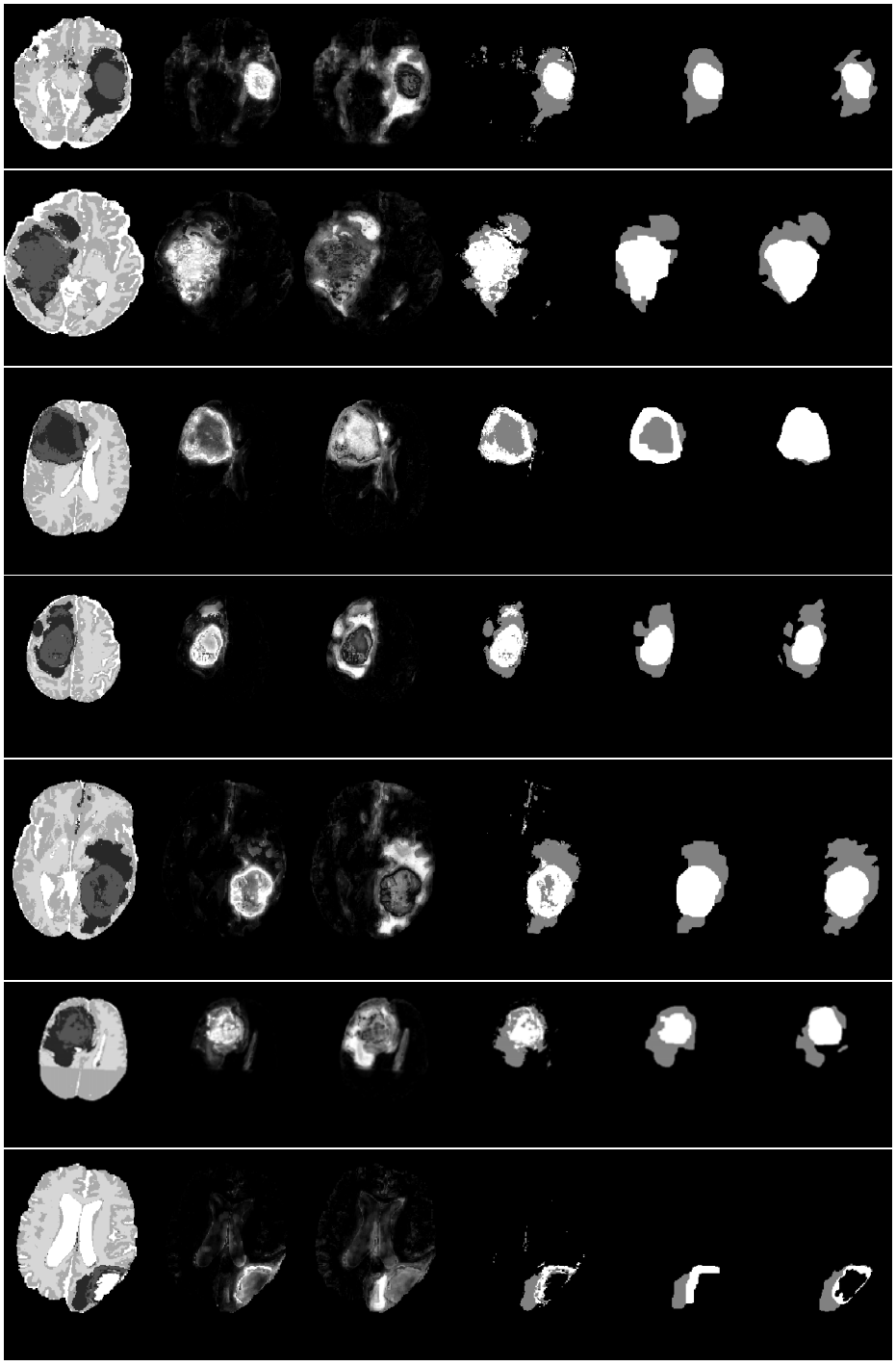

**Fig. 1.** Representative results of the tumor segmentation. Shown are the maximum-a-posteriori (MAP) estimates as obtained from the random forest for normal and tumor classe), the probabilities for core and edema (column 2,3), the MAP estimates of the two tumor classes before and after spatial smoothing (column 4,5), and the ground truth (column 6). The examples show that expert annotation may be disputable in some cases.

| ID | Dice1 | Sens1 | Spec1 | Dice2 | Sens2 | Spec2 |
|---|---|---|---|---|---|---|
| BRATS_HG0027 | 0.735 | 0.823 | 0.995 | 0.822 | 0.898 | 0.997 |
| BRATS_HG0026 | 0.738 | 0.758 | 0.997 | 0.412 | 0.401 | 0.998 |
| BRATS_HG0025 | 0.641 | 0.934 | 0.992 | 0.06 | 0.031 | 1 |
| BRATS_HG0024 | 0.7 | 0.834 | 0.997 | 0.896 | 0.982 | 0.999 |
| BRATS_HG0022 | 0.779 | 0.806 | 0.998 | 0.821 | 0.729 | 1 |
| BRATS_HG0015 | 0.8 | 0.82 | 0.996 | 0.894 | 0.879 | 0.999 |
| BRATS_HG0014 | 0.327 | 0.476 | 0.994 | 0.761 | 0.696 | 0.998 |
| BRATS_HG0013 | 0.7 | 0.661 | 1 | 0.887 | 0.985 | 1 |
| BRATS_HG0012 | 0.629 | 0.704 | 0.999 | 0 | 0 | 1 |
| BRATS_HG0011 | 0.808 | 0.763 | 0.998 | 0.9 | 0.889 | 0.999 |
| BRATS_HG0010 | 0.664 | 0.788 | 0.999 | 0.836 | 0.879 | 1 |
| BRATS_HG0009 | 0.833 | 0.822 | 0.997 | 0.749 | 0.604 | 1 |
| BRATS_HG0008 | 0.784 | 0.679 | 0.999 | 0.917 | 0.979 | 0.998 |
| BRATS_HG0007 | 0.644 | 0.508 | 0.999 | 0.838 | 0.942 | 0.999 |
| BRATS_HG0006 | 0.7 | 0.795 | 0.994 | 0.793 | 0.731 | 0.999 |
| mean | 0.699 | 0.745 | 0.997 | 0.706 | 0.708 | 0.999 |
| median | 0.7 | 0.788 | 0.997 | 0.822 | 0.879 | 0.999 |

| ID | Dice1 | Sens1 | Spec1 | Dice2 | Sens2 | Spec2 |
|---|---|---|---|---|---|---|
| BRATS_LG0015 | 0.402 | 0.751 | 0.997 | 0 | 0 | 1 |
| BRATS_LG0014 | 0.405 | 0.605 | 0.999 | 0 | 0 | 1 |
| BRATS_LG0013 | 0.29 | 0.492 | 0.996 | 0.164 | 0.089 | 1 |
| BRATS_LG0012 | 0.424 | 0.94 | 0.996 | 0 | 0 | 1 |
| BRATS_LG0011 | 0.3 | 0.908 | 0.994 | 0 | 0 | 1 |
| BRATS_LG0008 | 0.419 | 0.53 | 0.999 | 0.521 | 0.397 | 1 |
| BRATS_LG0006 | 0.767 | 0.992 | 0.998 | 0.788 | 0.74 | 1 |
| BRATS_LG0004 | 0.813 | 0.898 | 0.998 | 0 | 0 | 1 |
| BRATS_LG0002 | 0.652 | 0.767 | 0.989 | 0.017 | 0.009 | 1 |
| BRATS_LG0001 | 0.454 | 0.552 | 0.999 | 0.843 | 0.915 | 0.999 |
| mean | 0.493 | 0.744 | 0.996 | 0.233 | 0.215 | 1 |
| median | 0.422 | 0.759 | 0.998 | 0.009 | 0.005 | 1 |

**Table 1.** Performance measures as returned by the online challenge tool (`challenge.kitware.com/midas/`)indicating Dice score, sensitivity and specificity (top: high-grade cases; bottom: low-grade cases). Class "1", with results shown in the left column, refers to the "edema" labels. Class "2", with results shown in the right column, refers to the "tumor core" labels (for both low and high grade cases). Note that this definition differs somewhat from the labels returned by the algorithm that only indicates T1gad hyper-intense regions as class 2, irrespectively of the grading (low/high) of the disease.

8. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE T Med Imaging **20** (2001) 677–688
9. Breiman, L.: Random forests. Mach Learn J **45** (2001) 5–32
10. Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.A., Ayache, N., Golland, P.: Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation. In: Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge). (2012) 7p