

Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features

Wei Wu · Albert Y. C. Chen · Liang Zhao · Jason J. Corso

Received: 28 January 2013 / Accepted: 3 July 2013 / Published online: 17 July 2013
© CARS 2013

Abstract

Purpose Detection and segmentation of a brain tumor such as glioblastoma multiforme (GBM) in magnetic resonance (MR) images are often challenging due to its intrinsically heterogeneous signal characteristics. A robust segmentation method for brain tumor MRI scans was developed and tested. **Methods** Simple thresholds and statistical methods are unable to adequately segment the various elements of the GBM, such as local contrast enhancement, necrosis, and edema. Most voxel-based methods cannot achieve satisfactory results in larger data sets, and the methods based on generative or discriminative models have intrinsic limitations during application, such as small sample set learning and transfer. A new method was developed to overcome these challenges. Multi-modal MR images are segmented into superpixels using algorithms to alleviate the sampling issue and to improve the sample representativeness. Next, features were extracted from the superpixels using multi-level Gabor wavelet filters. Based on the features, a support vector machine (SVM) model and an affinity metric model for tumors were trained to overcome the limitations of previous generative models. Based on the output of the SVM and spatial affinity models, conditional random fields theory was applied to segment the tumor in

a maximum a posteriori fashion given the smoothness prior defined by our affinity model. Finally, labeling noise was removed using “structural knowledge” such as the symmetrical and continuous characteristics of the tumor in spatial domain.

Results The system was evaluated with 20 GBM cases and the BraTS challenge data set. Dice coefficients were computed, and the results were highly consistent with those reported by Zikic et al. (MICCAI 2012, Lecture notes in computer science, vol 7512, pp 369–376, 2012).

Conclusion A brain tumor segmentation method using model-aware affinity demonstrates comparable performance with other state-of-the art algorithms.

Keywords Brain tumor segmentation · Model-aware affinity · SVM · CRF · Superpixels

Introduction

Automatic brain tumor segmentation in magnetic resonance (MR) images has been an active research area for over a decade. Correct quantification of brain tumor size and growth is paramount to accurate prognosis and treatment planning; yet, manual delineation of brain tumor in dense MR images is costly and prone to error [1]. However, automatic detection and quantification of brain tumor are very difficult. MR images are often corrupted by factors, such as poor radio frequency (RF) coil uniformity, inhomogeneous static field, etc. [2]. Although methods for intensity standardization and inhomogeneity correction are widely known [2–4], valuable information is still lost in this process [4]. Furthermore, brain tumors are highly varied in size, shape, and appearance [5], as shown in Fig. 1. It is, for instance, difficult or implausible to segment a GBM by a simple threshold [6].

W. Wu · A. Y. C. Chen · L. Zhao · J. J. Corso
Department of Computer Science and Engineering,
SUNY at Buffalo, Buffalo, NY, USA

W. Wu
School of Information Engineering, Wuhan University
of Technology, Wuhan, China

W. Wu (✉)
Key Laboratory of Fiber Optic Sensing Technology
and Information Processing, Wuhan University of Technology,
Wuhan, China
e-mail: hustwuwei@gmail.com

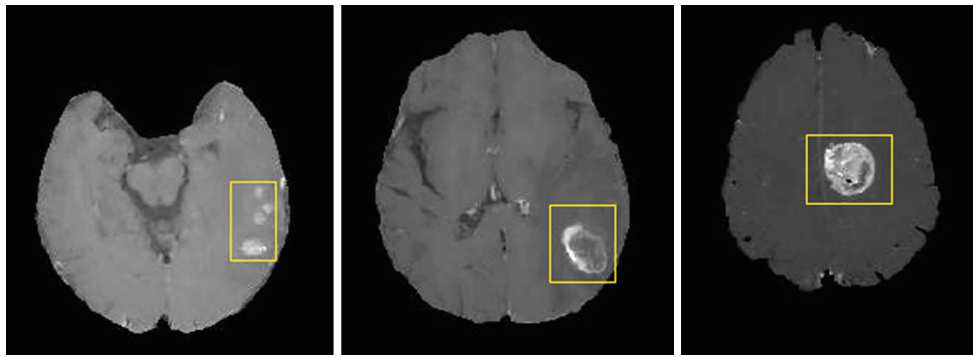


Fig. 1 Examples of tumor in three slices from different samples (T1CE channel)

Voxel-based methods have seen wide use for brain tumor segmentation. Philips et al. [7] give an early proof-of-concept fuzzy clustering by operating on the raw multi-sequence data. Clark et al. [8] integrate knowledge-based techniques and multi-spectral histogram analysis. Fletcher-Heath et al. [9] take a knowledge-based fuzzy clustering approach followed by 3D-connected components to build the tumor shape. Prastawa et al. [6] also present a knowledge-based detection algorithm that incorporates tumor-specific learned voxel intensity distributions. Voxel-based statistical classification methods are also used [10–12]. However, these methods are limited by their high degree of locality, since they do not take global context into account. As a result, their performance on larger data sets remains unsatisfactory.

Methods based on generative models have also been applied to tumor segmentation. Guillemaud et al. and Wells III et al. use a Gaussian Mixture Model (GMM) method [13, 14]. Corso et al. [15] present a brain tumor multi-level detection algorithm with integrated Bayesian model classification. The statistical model is also a GMM. Voxel-based generative models have intrinsic limitations: It is hard to construct an accurate model when the number of samples is insufficient, and spatial information is not taken into account, for all the data points are considered to be independent samples drawn from the sample population.

Recent advances in machine learning have had an impact in brain tumor segmentation: Methods based on discriminative classifiers have become a strong emphasis for tumor segmentation. For example, Zhou et al. [16] use SVMs, Xuan et al. [17] and Corso et al. [18] use the Adaboost algorithm [19], Dettling et al. [20] adopt the random forest algorithm [21], and Zikic et al. [22] use decision forests based on models integrating context-aware spatial features and GMM. In practice, however, it remains a challenge to learn a single, satisfactory classifier. For example, if every single pixel is considered as a training sample for the classifier, the amount of data in the training set is too huge to be applied directly. Therefore, how one should sample and train becomes a key problem, and it strongly impacts the classifier performance [23].

Note also that most GBM brain tumors are heterogeneous and that the learning problem is inherently imbalanced (most data are still normal brain) [24], both of which further increase the difficulty of the problem.

To overcome the limitations of strictly voxel-based methods, Markov and conditional random fields (MRFs) have also been applied to brain tumor segmentation [25]. They allow the classification of one voxel to depend on the labels (the classification) of neighboring voxels. Zhang et al. [26] realize segmentation through a hidden MRF and expectation-maximization (EM) algorithm. Corso et al. [18] use a hierarchical minimization algorithm to segment voxels in a conditional random field (CRF) framework with features learned by Adaboost. Lee et al. [27] use a context-sensitive discriminative random fields model. They incorporate a set of knowledge-based features coupled with support vector machines in a voxel-level CRF framework to perform the segmentation and classification. The CRF-based methods have summarily outperformed earlier voxel classifier-based methods. However, to the best of our knowledge, all previous CRF-based methods for brain tumor detection and segmentation have a weak model on the binary potential relating to pairs of labels: Most use the Ising or Potts model, and Lee et al. [27] use a feature difference model. These simple affinity models fail to capture the potential rich relational features that can be modeled on neighboring pixels.

In this paper, we improve on the CRF-based methods to incorporate a model-aware affinity metric: The binary affinity relating two neighboring pixels is based on the likelihood of the two pixels being a particular tissue type, and hence, it varies as a function of the two pixels' labels. The model-aware affinity was first proposed in our earlier work [15], but not used within the context of the CRF framework. Based on the affinity model, the relationship between neighboring nodes becomes more explicit and interpretable, for they are learned from the training set directly. In addition to adopting the model-aware affinity on the CRF edges, we also incorporate a superpixel-level affinity function to regularize the features over many pixels.

We use a state-of-the-art superpixel oversegmentation algorithm (graph-based image segmentation [28]) to segment the multi-channel MR images, alleviating the sampling issue and improving the sample representativeness. Although the superpixel oversegmentation algorithm can be applied in three dimensions volumetrically [29], we strictly use two-dimensional segmentation in this paper, to account for data sets with limited resolution in one or more dimensions. Then, multi-level Gabor wavelet filters are adopted to extract features from the superpixels. Based on the features, an SVM classifier and a model-aware affinity model are constructed and trained for tumors. With the output of the SVM (scores), we realize the tumor detection and segmentation with a CRF that uses a model-aware affinity for the binary term. Finally, we incorporate a set of denoising routines on the CRF output based on the structural properties of the brain and the tumor. Our extensive experiments demonstrate that the proposed method systematically outperforms both the original model-aware affinity method due to Corso et al. [15] and the CRF method of Lee et al. [27].

Mathematical background

This section introduces the model-aware affinity model and the CRF model for brain tumor detection and segmentation. The methods for parameter learning of these models are also given.

Model-aware affinity

The input data induce a graph with each node in the graph denoted $u, v \in S$ where S is the full spatial coordinate system (the lattice) and s_μ is the set of properties (intensity or texture) at node μ . In a GB representation for segmentation, we consider graph nodes for each voxel in the image and annotated graph edges in which the weight represents a so-called *affinity* of two neighboring nodes. Intuitively, affinity is the preference of the two nodes to be grouped together into a single segment; it is commonly but does not need to be defined directly by pixel feature similarity. The affinity is denoted by $\omega_{\mu\nu}$ for connected nodes (i.e., μ and ν are neighbors in the graph). We consider affinity functions of the following form as in [15]:

$$\omega_{\mu\nu} = \exp(-D(s_\mu, s_\nu; \theta)), \tag{1}$$

In this work, the GBM MR image data set we use has four channels: T1 weighted pre-contrast (T1) and post-contrast (T1CE), FLAIR and T2 weighted. D is a non-negative distance measure, and $\theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ is the set of learned weights for the four channels.

There are three kinds of “objects” or classes of tissue in the MR images at hand: normal brain tissue, tumor, and edema.

The classes are denoted as random variables $m : m_\mu, m_\nu$ are the model labels at nodes μ and ν , respectively. We define a model-aware affinity function (see [15] for a full derivation of the model-aware affinity, which is based on a Bayesian interpretation of the affinity as a binary random variable):

$$\omega_{\mu\nu}(m_\mu, m_\nu) = \exp(-D(s_\mu, s_\nu; \theta[m_\mu, m_\nu])). \tag{2}$$

In order to promote efficient calculation, we use the following functional form of $D(\cdot)$:

$$\omega_{\mu\nu}(m_\mu, m_\nu) = \exp\left(-\sum_c \theta_{m_\mu m_\nu}^c |s_\mu^c - s_\nu^c|\right), \tag{3}$$

where superscript c indicates vector θ at channel c . The class-dependent coefficients are learned from the labeled training data by constrained optimization. Intuitively, the affinity between two nodes of same class should be near 1, and near 0 for different classes. Thus, we learn the coefficients by optimizing the following functions under the constraint that the coefficients sum to one for each class pair:

$$\begin{aligned} & \sum_c \theta_{m_\mu, m_\nu}^c = 1, \forall \{m_\mu, m_\nu\}: \\ & \theta_{m_\mu^*, m_\nu^*} = \begin{cases} \arg \max_{\theta} \sum_{\substack{u : m_\mu = m_\mu^* \\ v : m_\nu = m_\nu^*}} \exp\left(-\sum_c \theta_{m_\mu, m_\nu}^c |s_\mu^c - s_\nu^c|\right), & \text{if } m_\mu^* = m_\nu^* \\ \arg \min_{\theta} \sum_{\substack{u : m_\mu = m_\mu^* \\ v : m_\nu = m_\nu^*}} \exp\left(-\sum_c \theta_{m_\mu, m_\nu}^c |s_\mu^c - s_\nu^c|\right), & \text{otherwise} \end{cases} \end{aligned} \tag{4}$$

To learn the parameters and obtain the optimal coefficients for each class pair, we first set an initial parameter based on stochastic sampling. Then, we adopt a steepest coordinate descent procedure to search for the best coefficients. The gradient of the function is estimated numerically at each iteration, and the single coordinate of optimal modification is pursued. When no further improvements can be made, the iterative procedure is terminated. Convergence to some local optimal point is guaranteed.

A new image I_{map} , which we call the affinity map of the tumor, can be formed using the parameter θ and the four channel original images based on the affinity function in Eq. (1). In order to avoid decay in our future steps, we redefine a new affinity relation between two nodes μ and node ν based on an affinity map of the tumor I_{map} .

$$w(\mu, \nu) = 1 - \frac{|I_{\text{map}}(\mu) - I_{\text{map}}(\nu)|}{\max(I_{\text{map}})}. \tag{5}$$

Figure 2 is an example of an affinity map of the tumor. Here, s_μ is a 4D vector whose values come from the intensities in the 4-channel image as in [15]. We notice that although there is high noise, the regions of the tumor have high intensity, implying the pixels in the regions have strong similarity and hence have a preference to be grouped (note this is the tumor

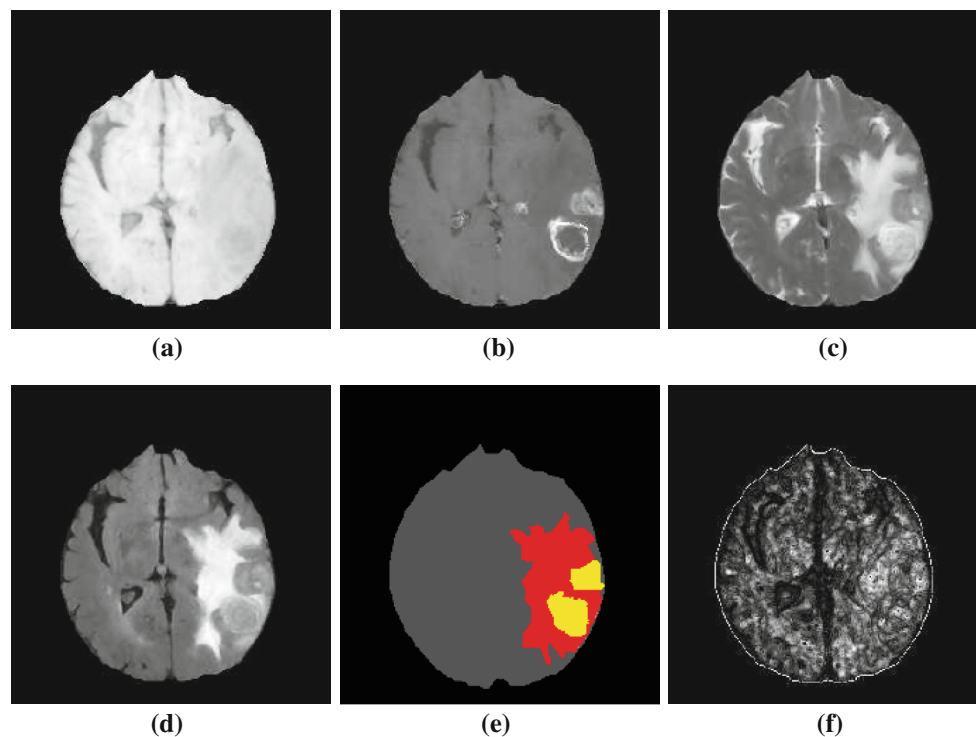


Fig. 2 An example of the affinity map of the tumor class. [The *yellow* region is edema, and the *red* region is tumor in image (e)]. **a** T1, **b** T1CE, **c** T2, **d** Flair, **e** the ground truth, **f** the affinity map of the tumor

class-specific affinity map), which is very important in our future process.

Conditional random fields

CRFs are a discriminatively trained MRFs. The underlying idea is that of defining a conditional probability distribution over label variables given a particular observation, rather than a joint distribution over both labels and observations. This subtle difference alleviates the need to model the distribution over the observations, which is important in medical imaging applications since tumors have complex shape and intensity distributions that are not easy to model and may not be appropriately modeled in factorized form. CRFs have been used to some degree in brain tumor imaging [27,30].

We use a standard pairwise CRF model:

$$p(m|x) = \frac{1}{Z} \exp \left(\sum_{\mu \in S} A_{\mu}(m_{\mu}, x_{\mu}) + \sum_{\mu \in S} \sum_{v \in N_{\mu}} I_{\mu v}(m_{\mu}, m_v, x_{\mu}) \right), \quad (6)$$

where A_{μ} is the association (observation matching) potential for modeling dependencies between the class label m_{μ} and the set of all observations μ . x_{μ} is the real-valued SVM

response on the pixel (or node) μ . N_{μ} is the neighborhood of pixels μ (a subset of the full spatial coordinate system S from above). $I_{\mu v}$ is the interaction (local-consistency) potential for modeling dependencies between the levels of neighboring elements. Z is the partition function: a normalization coefficient (sums over possible labels).

While Eq. (6) holds for a set of multiple labels m_{μ} , herein, we only use it for 2 classes with $m_{\mu} = 1$ and $m_{\mu} = -1$. We use an SVM to represent the association potential [27,31] (we realize many representations are plausible—see the related work discussed in the Introduction Section—but we choose the SVM based on its generalization characteristics):

$$A_{\mu}(m_{\mu}, x_{\mu}) = \frac{1}{1 + \exp(B \times x_{\mu} + C)}, \quad (7)$$

where the parameters B and C are estimated from the training data represented as pairs (x_{μ}, t_{μ}) , and t_{μ} denotes a related probability that $m_{\mu} = 1$, represented as the relaxed probabilities: $t_u = \frac{N_+ + 1}{N_+ + 2}$ if $m_{\mu} = 1$, and $t_u = \frac{1}{N_- + 2}$ if $m_{\mu} = -1$, where N_+ and N_- are the number of positive and negative class instances.

Platt et al. [31] use a Levenberg–Marquardt approach that tries setting C to guarantee that the Hessian approximation is invertible. Using these training instances, we can solve the following optimization problem to estimate parameters B and C (see [27] for a detailed solution),

$$\min - \sum_{x_\mu \in \Omega} [t_\mu(B \times x_\mu + C) + \log(1 + \exp(-B \times x_\mu - C))]. \tag{8}$$

Ω denotes the set of training instances.

We adopt the following pairwise local-consistency potential:

$$I_{\mu\nu}(m_\mu, m_\nu, x_\mu) = m_\mu m_\nu (\gamma \phi_{\mu\nu}(x_\mu)). \tag{9}$$

The coefficient $\phi_{\mu\nu}(x_\mu)$ expresses the relation for sites μ and ν . The specification of this functional coefficient is one main advantage of our paper: The inhomogeneity characteristics of MR images and tumor images, specifically, are high. It is, hence, improper to define $\phi_{\mu\nu}(x_\mu)$ as constant function, as has been defined in early methods using MRF (or CRF) in MR images for brain tumor detection, such as an Ising model in [30]. Lee et al. [27] define $\phi_{\mu\nu}(x_\mu)$ as the pixel-level feature vector difference, to better reflect the actual input to the SVM classifiers. Their setup remains unable to capture the full variation of local consistency, as we show in our experiments. Since two neighboring pixels’ feature vectors are extracted from two highly overlapping regions, their local-consistency potential $\phi_{\mu\nu}(x_\mu)$ has little effect on the random field model. Furthermore, their model is subject to and tends to propagate the noise introduced at the feature extraction level, as opposed to the typical settings where the local-consistency potential acts as contextual priors that help to overcome the noise introduced in the feature extraction and classifier training process.

In contrast, we use a generative model (model-aware affinity) to represent the relation between neighboring pixels, which is more explicit and interpretable, for it is learned from the training set directly. Equation (5) is used to calculate $\phi_{\mu\nu}(x_\mu)$. Thus, the spatial affinity model has been directly integrated into the CRF, and our results demonstrate the importance of such an adaptive binary term in the CRF over the earlier, simpler, fixed terms or classifier-based terms.

The coefficient γ is the vector of local-consistency parameters to be learned. We estimate it from the l training pixels

from each of the K training images using pseudo-likelihood (see [27] for the full solution):

$$\hat{\gamma} = \arg \max_{\gamma} \left\{ \sum_{k=1}^K \sum_{\mu \in \Omega} \{p(m_\mu | x_\mu)^n + \sum_{v \in N_\mu} m_\mu m_v (\gamma \phi_{\mu\nu}(x_\mu)) - \log(z_\mu^k)\} - \frac{1}{2\tau} \gamma^T \gamma \right\} \tag{10}$$

where z_μ^k is a partition function for each site μ in image k and τ is a regularizing constant. Keeping the observation matching constant, the optimal local-consistency parameters γ can be found by gradient descent.

Our approach for brain tumor segmentation

Our framework for brain tumor detection is shown in Fig. 3. For the training data, the multi-channel MR images are segmented into many superpixels using [28]. Then, the features are extracted from the superpixels using multi-level Gabor wavelet filters. Based on the features and labels of the superpixels, the SVM model is trained and an affinity model for tumor is learned from the training data set. The parameters of the CRF are also computed in the stage. For testing data, the output of the SVM and the spatial affinity model is integrated into the CRFs. After the brain tumors are segmented by the CRF, false positives are effectively removed in a structural denoising process. In the subsequent subsections, we explain these system parts in detail.

Superpixel segmentation

Many existing algorithms in medical imaging use pixels as the underlying representation [5, 6, 11, 12]. Pixels, however, are not a natural representation of visual scenes. They are rather an “artifact” of the MR imaging process, similarly as in natural images [32]. Superpixels—regions in an over-segmentation of the image—would be more natural and pre-

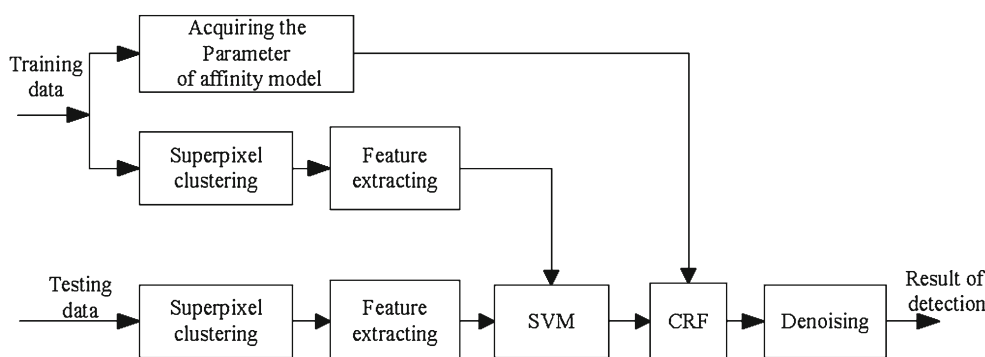


Fig. 3 Framework of brain tumor detection

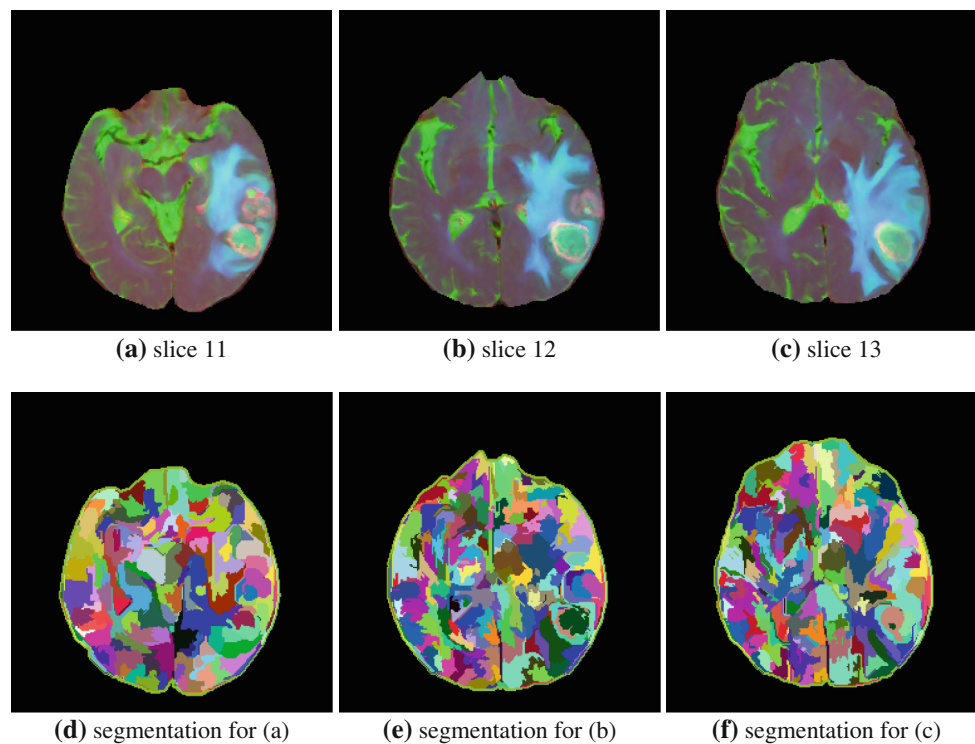


Fig. 4 An example segmentation based on GB algorithm. For the visual convenience, we use a pseudocolor rendering of each MR slice. The pseudocolor is composed of channel T1CE, T2, and Flair. **d–f** are

the segmentation results of **a**, **b**, and **d**, where each superpixel is randomly and uniquely colored

sumably lead to more efficient processing. Superpixels have many desirable properties:

- (a) Computational efficiency: Superpixels reduce the complexity of images from hundreds of thousands of pixels to only a few hundred superpixels.
- (b) Representational efficiency: Pairwise constraints between superpixels, while only for adjacent pixels on the pixel grid, can now model much longer-range interactions between superpixels.
- (c) The superpixels are perceptually meaningful: Each superpixel is a perceptually consistent unit; all pixels in a superpixel are most likely uniform or at least homogeneous.
- (d) Completeness: Because superpixels result from an over-segmentation, many structures in the image are conserved. There is little loss in moving from the pixel grid to the superpixel map.

There are many methods for obtaining superpixels. We adopt the graph-based (GB) algorithm because it has good affinity in the superpixels. Although the shape of its superpixels is slightly irregular, it has better precision than the other superpixels method in the recent benchmark study [28,29]. The basic idea of the algorithm is based on converting the image into a graph and then manipulating the graph to segment the

image. In the process of converting the image into a graph, each pixel has a corresponding vertex and each pixel vertex has an edge to its neighbors (inherited from the pixel lattice). The weight of edges between pixel vertices is determined by a function that promotes homogeneity, and the graph is segmented to minimize the energy function defined by the edge weights. The algorithm can be applied for ND images.

Figure 4 is an example segmentation based on the GB superpixels algorithm. We apply the algorithm to take the MR image as a set of 2D images with 4 channels (recall our transverse resolution is too low to consider it as a true 3D volume). Every pixel is a 4D vector that contains the value of the corresponding position from images T1, T1CE, Flair, and T2. The four channel values are normalized to a common scale. Then, we can apply the GB superpixels algorithm to obtain superpixel regions. Image (a), (b), and (c) are three neighboring slices (slice numbers 11, 12, and 13, respectively) from the B1 data set (the data set is explained later in section “Experiments”). From the figure, we can see that the superpixels effectively represent homogeneous regions in the MR images without violating the tumor/edema boundaries.

Feature extraction

We use a Gabor wavelet transform to extract features [33,34] of the brain tumor. Among various wavelet bases, Gabor

functions provide the optimal resolution in both the time (spatial) and frequency domains, and the Gabor wavelet transform is the preferred basis to extract local features for several reasons:

- Biological motivation: The simple cells of the visual cortex of mammalian brains are best modeled as a family of self-similar 2D Gabor wavelets [34].
- Mathematical and empirical motivation: The Gabor wavelet transform has both the multi-resolution and multi-orientation properties and is optimal for measuring local spatial frequencies. It has been found to yield a distortion tolerance space for pattern recognition tasks [34].

In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave as follows (we apply the superpixel segmentation at each slice separately due to limited transverse resolution and hence apply the feature extraction in the planes too):

$$G(x, y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \exp(j\omega(x \cos \theta + y \sin \theta)) \quad (11)$$

where σ is the standard deviation of the Gaussian function in the x - and y -directions, and $j\omega$ denotes the spatial frequency. A family of Gabor kernels can be obtained from Eq. (11) by selecting different center frequencies and orientations. These kernels are used to extract features from an image.

Due to the uncertain brain tumor size, we use 5 scales of Gabor wavelets in the experiments, each with 8 directions, as shown in Fig. 5. After filtering by the Gabor wavelet, we calculate the mean value of response for every superpixel, and by adding the mean intensity (4 channels), each superpixel has 164 ($5 \times 8 \times 4 + 4$) features.

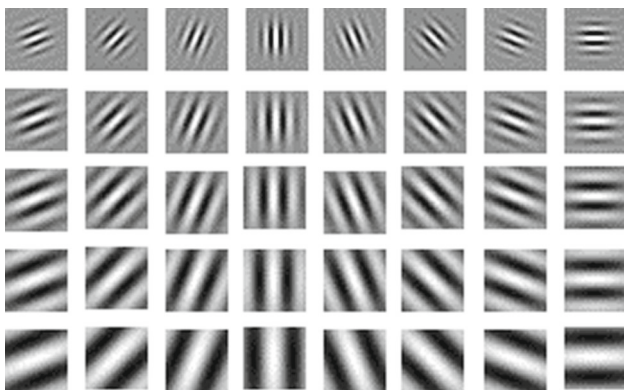


Fig. 5 Gabor filters: 5 scales with 8 directions

Model formulation, solution, and application

SVM training and application

Due to the advantages of superpixels, it is possible to train a classifier without paying excessive attention to data sampling issues while obtaining comparatively better precision (avoiding improper sampling) and less calculation (decreasing the amount of data) than training the classifier directly on pixel samples.

For training, every superpixel is labeled based on the ground truth. We use the labels and features of the superpixels to train the SVM model, but the common challenge of data imbalance is present [24]. Most of the ground truth data are non-tumor; only a small amount is tumor. We use hard negative mining to handle this issue. A “hard negative” is a false positive detection, meaning that a classifier marked a non-tumor sample as tumor. The SVM’s learning can be refined by iteratively incorporating these hard negatives into the training set and retraining the SVM.

The hard negative mining proceeds as follows. The SVM is first trained on random negative samples, establishing a baseline. Then, the trained SVM is evaluated on a set of strictly non-tumor data, so that any positives it returns are guaranteed to be false positives. A subset of the collected false positives is then incorporated into the negative training samples, and the SVM is retrained. This approach improves the SVM at each iteration, since difficult samples that it has misclassified are disambiguated and then explicitly trained into it. In practice on our data, the SVM’s overall accuracy is improved by 1%. Here, overall accuracy of SVM is $(T_p + T_n)/(T_p + T_n + F_n + F_p)$, T_p is the true positive and T_n is the true negative, F_p is the false positive and F_n is the false negative.

CRF training and application

The results of the SVM are based on superpixels. Although minor, it is inevitable there is some precision loss based on superpixels. In order to recover the loss, we use the CRF model based on pixel affinity. We use Eq. (4) to study the spatial relation between the pixels. Equation (8) is applied to acquire the transform parameters from the response of SVM to the association potential. Equation (10) is applied to confirm the coefficient of CRFs.

For testing data, we use Eqs. (5) and (7) to compute the association potential of pixels and associate it with the relevant edges in the CRF (the SVM’s scores of the pixels are assigned the value of superpixel to which they belong). The CRF is based on a standard pixel framework. For every pixel, the relation between it and its neighbor pixels is calculated by Eq. (9). In this paper, we strictly use a 2D CRF because the transverse resolution of our data is low, but we note that

a direct extension of the approach to a 3D CRF is plausible. We, hence, adopt a 4-pixel neighbor. For convenience of application, we express the relations for all pixels using two edge maps. One is the horizontal relation map, the other is vertical relation map. Then, the CRF is applied to smooth and segment the tumor regions [35–39].

Structural denoising

The candidate regions (the results of output CRF) may contain false positive pixels. We remove some false positive pixels according to structural properties of the brain and the tumor: symmetry and continuity of the brain/tumor. Similar knowledge-based routines have been proposed in previous brain tumor research [6,8,12], but primarily as the main tumor segmentation methodology and not as a structural denoising post-processing step.

Denoising based on symmetrical characteristic

The position of most brain tumors is random. Most of them are not symmetrical if they are measured from the axis of the brain [40], but most normal tissues are approximately symmetrical about the axis of the brain. Therefore, we can use the symmetrical characteristic to remove some false positives. The midline axis of the brain is very important, even in the case of midline shift from brain tumor [6].

We propose an algorithm that can effectively detect the symmetry plane in our data. The algorithm is based on the theory of dynamics and assumes the symmetry axis is a rigid body with the minimal value of inertia. According to the theory of dynamics, the moment of inertia is defined as: $I = \sum r^2 dm$.

In the image, every pixel is taken as a small rigid object. If the function of symmetry axis is $y = kx + b$, then we can get a moment of inertia of a pixel as $I_i = \sum_{i=1}^N \alpha D_i^2$, where $D_i = \frac{|y_i - kx_i - b|}{\sqrt{1+k^2}}$, α is a constant set to 1. Now, we solve the parameter k and b according to the condition of minimizing the value I_i ,

$$\begin{cases} \frac{\partial I_i}{\partial b} = \frac{1}{1+k^2} \sum_{i=1}^N (b + kx_i - y_i) = 0 \\ \frac{\partial I_i}{\partial k} = \frac{\sum_{i=1}^N (b+kx_i - y_i)}{(1+k^2)^2} \sum_{i=1}^N (x_i - kb + ky_i) = 0 \end{cases} \quad (12)$$

We get $k_{1,2} = \frac{-N \pm \sqrt{4M^2 + N^2}}{2M}$, in which we choose an appropriate value for k according to the prior information, where $M = \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})$, $N = \sum_{i=1}^N (x_i - \bar{X})^2 + (y_i - \bar{Y})^2$, $\bar{X} = \sum_{i=1}^N x_i / N$, $\bar{Y} = \sum_{i=1}^N y_i / N$, $b = \bar{Y} - k\bar{X}$.

After getting the symmetry axis, we make a symmetrical transform about the symmetry axis for the pixels in the image. If the coordinate of a pixel in image is (x_0, y_0) , the

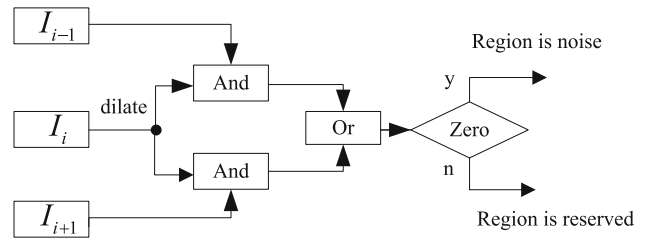


Fig. 6 Flow diagram for the continuity-based denoising filter

symmetrical coordinate of the pixels is (x_1, y_1) :

$$\begin{cases} x_1 = 2x_c - x_0 \\ y_1 = 2y_c - y_0 \end{cases}, \quad (13)$$

where $x' = \frac{x_0 + (y_0 - b)k}{k^2 + 1}$, and $y' = kx_c + b$. Since it is unlikely that two symmetrical points (x_0, y_0) and (x_1, y_1) are both classified as tumor, we treat such occurrences as noise.

Denoising based on continuity

Tumors, unlike random noise, occupy a continuous volume in the MR images. A 3D CRF could remove the discontinuous noise in transverse plane if consistency in z -direction is ensured at some point. When the transverse resolution is too low—as in our data—we design a continuity filter to remove the noise, as shown in Fig. 6. If a tumor is detected in an image, we expect to find tumor in the previous or next slice accordingly.

The image is dilated first, and then, it is intersected with the two adjacent slices (previous and next). Finally, we make a decision to remove the false positive regions: If both results are zero, the region is treated as noise; otherwise, the region is preserved.

Experiments

For the ease of comparison, we first use the same data set as in [15], and the same twofold cross-validation testing approach. The data set consists of 20 expert-annotated GBM studies. The data set is pre-processed through the following pipeline: (1) spatial registration, (2) noise removal, (3) skull removal, and (4) intensity standardization. Following [15], the data are split into two groups, ‘A’ and ‘B’, and each experiment is performed on twofold cross-validation (first train on ‘A’ and test on ‘B’ and then train on ‘B’ and test on ‘A’).

The experiments that we have conducted evaluate both the various aspects of our method as well as make a thorough comparison with the state of the art. In the results, we denote the method of Corso et al. [15] as “Model-Based SWA” (SWA, segmentation by weighted aggrega-

Table 1 Comparison of SVM's accuracy of brain tumor (%)

	Method based on pixels				Method based on superpixels			
	OA	J	P	R	OA	J	P	R
Train data (set A)	95.8	45.2	97.3	45.8	99.4	78.7	88.5	87.7
Test data (set B)	96.6	38.0	88.1	40.1	97.5	52.3	66.6	70.9
Train data (set B)	97.0	51.9	97.7	52.6	99.5	75.8	91.4	81.6
Test data (set A)	95.0	40.1	88.3	42.3	97.4	46.3	58.5	69.0

In here, "OA" is overall accuracy score, "J" is Jaccard score, "P" is precision score, and "R" is recall score. Bold values indicate the best results

tion algorithm), the CRF method of Lee et al. [27] as "SVRF", the SVM classifier with a simple Ising model as "SVM", and our proposal CRF with SVM and model-aware affinity as "SVM + AM". SVM based on both superpixels and AM (affinity metric) are the main novelties in this paper.

Superpixels versus pixels

Our first experiment directly compares the use of superpixels against the raw pixels, with the SVM classifier. The comparison of the SVM's accuracy using methods based on pixels and superpixels is shown in Table 1. Here, overall accuracy of SVM is $(T_p + T_n)/(T_p + T_n + F_n + F_p)$, the Jaccard score is $T_p/(T_p + F_p + F_n)$, the precision score is $T_p/(T_p + F_n)$, the recall score is $T_p/(T_p + F_p)$, T_p is the true positive and T_n is the true negative, and F_p is the false positive and F_n is the false negative.

From Table 1, we find that methods using superpixels have better overall accuracy. It is easy to explain: There is noise in the MR images. If we use an SVM based on pixels, the training data are too large to be applied directly, so it must be sampled to train a model. Both of these two reasons cause a loss of overall accuracy. By adopting the superpixels, much of this pixel noise is smoothed during the segmentation of superpixels, and the number of training samples is also decreased. We can gather more effective samples to train the model, and it is even possible to load all the training samples (for this data set), therefore, yielding models with higher precision. A point worth mentioning is that the size of superpixels must be limited (i.e., one can over-segment too much). In the paper, the merging threshold is set to 5.

We also find the results of SVM based on superpixels have better performance than those based on pixels in the Jaccard score and the recall score. The method based on superpixels achieves a better balance between precision and recall than that based on pixels. This attribute is vital to the upcoming steps. Although pixel-based SVMs have good precision, the low recall due to the excessive amount of noise causes irreversible errors on the final output. Although the CRF model can smooth noise, if the noise is high, it hardly can be removed.

Evaluation of performance against state of the art

Figure 7 shows 6 slices segmentation from the test set, where each column shows a different test case. The first row is the pseudocolor displays of slices (pseudocolor is composed of channel T1CE, T2, and Flair, for the MR image in channel T1 is relatively smooth). The second row is the ground truth. In here, the red regions indicate edema and the yellow regions indicate tumor. The process of edema segmentation is same as that of brain tumor. The third row is the result of Ising model using SVM based on superpixels. The fourth row is the output of SVRF (Lee et al. in [27]), and the fifth row is the output from our algorithm (SVM + AM). For fair comparison, the features in method of SVRF are the same as our method. The affinity method of our algorithm is same as that of SWA method.

We can find that the results of our algorithm are closer to ground truth than those of the other two methods, in particular the detail along the boundaries, because the spatial information of tumor is integrated.

In Table 2, we show the Jaccard results taking a weighted averaged over the set. The results demonstrate that our algorithm has the best performance. Twofold cross-validation also confirms the conclusion. Notice the positive correlation of stronger spatial prior models with the SVM algorithm's performance. The simple spatial prior model used in the SVRF, which does not taken into account whether two pixels belong to the same tumor region or not, improves the SVM performance by 8%. Our spatial prior model, which penalizes the spatial discontinuity using model-level constraints, improves the SVM result by 12% while improving the SVRF result by 4%.

In Table 3, we give quantitative Jaccard scores for specific cases B1–B10 as in [15]. It is worth noting that despite our overall better performance, the simple GMM model used in [15] yields slightly better results on the individual samples B1–B4. This is due to the tumor region in those samples better satisfying the GMM assumption, therefore allowing a simpler generative model to outperform our more complex discriminative model.

From Tables 2 and 3, we also show that our algorithm has better performance than that in [15,27] for edema. The process for edema, in our method, is the same as that of tumor.

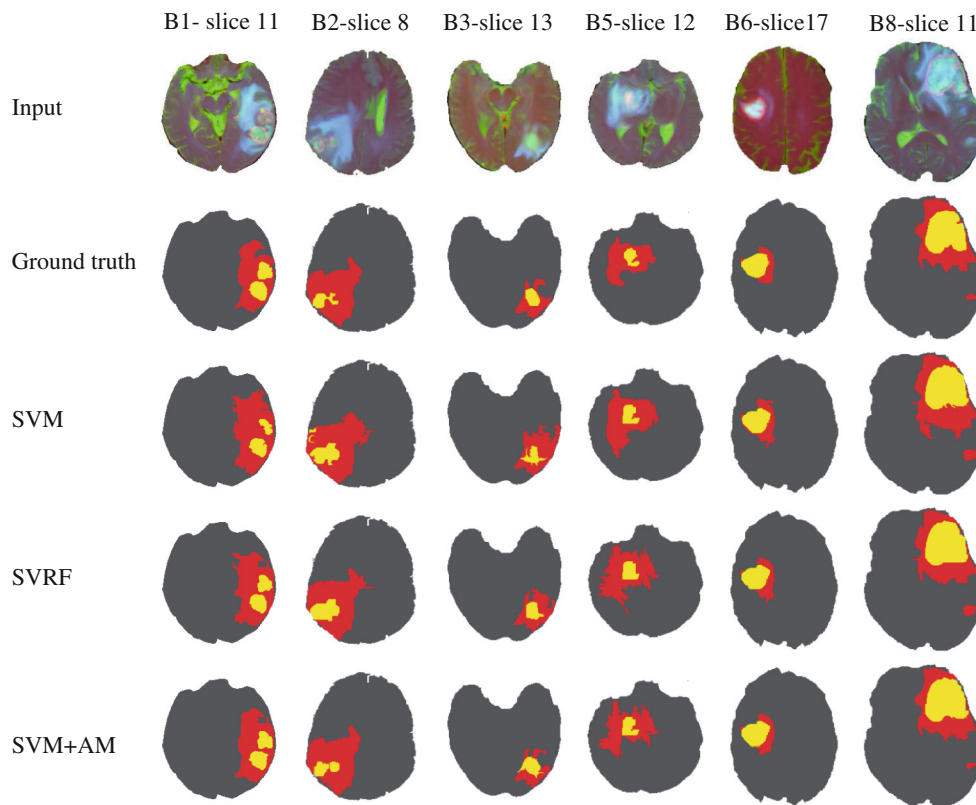


Fig. 7 Results of some slices from the test set. See the description in text

Table 2 Statistics of the result of (Jaccard %)

Algorithm	Tumor		Edema					
	Train	Test	Twofold cross-validation		Train	Test	Twofold cross-validation	
			Train	Test			Train	Test
Model-Based SWA	69	69	68	55	63	62	65	54
SVM	80	55	79	48	65	58	66	57
SVRF	81	64	80	56	68	61	68	62
SVM+AM	85	73	87	60	71	66	72	64

Bold values indicate the best results

The segmentation is done based on a CRF model for edema, including the affinity model and SVM for edema.

Importance of structural denoising

We study the proposed structural denoising for removing false positives in a post-processing step. Figure 8 is an example of denoising based on the symmetrical characteristics. Image (a) and (b) are the pseudocolor displays of slices 7 and 8 from B9 test set (pseudocolor is composed of channel T1CE, T2 and Flair, for the MR image in channel T1 is relatively smooth). Image (c) and (d) are the outputs of slice 7 and 8 based on CRF. *The two slices are normal tissue with no trace of edema or tumor.* So the yellow regions in

image (c) and (d) are both false positives for wrongly taking the symmetrical normal tissue as tumor. This false positive is effectively removed with our symmetrical characteristic-based noise removal process.

The method of denoising based on continuity is also effective. Some isolated noise is removed in 8 slices according to our statistics. As with many denoising algorithms, the improvement to the Jaccard score is minimal; yet, the qualitative improvement is easily noticeable to the human.

Failure modes

As shown in Table 3, our algorithm cannot get a satisfactory result in B7 test set, and the Jaccard score is worse than that

Table 3 Quantitative Jaccard score of Algorithm for tumor and edema (Jaccard %)

Data	Tumor				Edema			
	Model-Based SWA	SVM	SVRF	SVM + AM	Model-Based SWA	SVM	SVRF	SVM + AM
B1	58	32	45	50	73	59	63	70
B2	73	62	57	64	65	56	58	64
B3	80	73	72	76	73	60	59	68
B4	78	49	69	67	83	75	76	84
B5	72	86	82	87	66	64	79	84
B6	37	83	84	85	13	25	47	53
B7	27	24	29	35	33	59	81	66
B8	85	50	82	86	73	70	71	80
B9	72	50	65	71	75	45	79	67
B10	82	89	89	91	66	65	80	83

Bold values indicate the best results

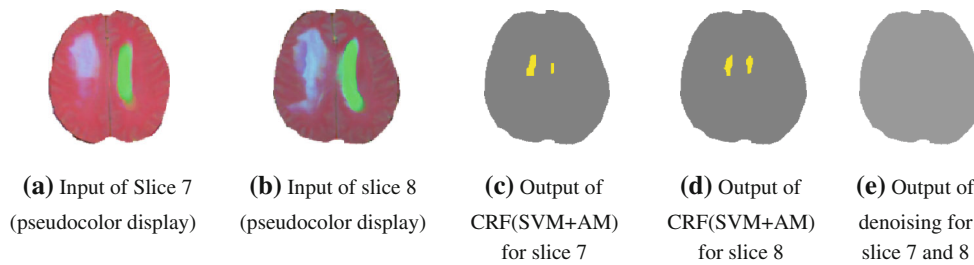


Fig. 8 An Example of denoising based on symmetrical characteristic. **a** and **b** are two continuous slices from a normal brain MR image, while **c** and **d** are false positives detections from the CRF step. Such types of

false positive are effectively removed with our noise removal process based on symmetrical characteristic

of the SWA algorithm in B1, B2, B3, and B4. Our observation is that the features we use, despite an improvement from the former methods, are still not optimal for the problem and cannot fully capture the diverse and complicated nature of the tumors. In order to overcome the problem in the future, we expect more effective features would improve the performance of the classifier. Furthermore, spatial features based on 3D superpixels can also be applied if the data set has have enough precision in the transverse plane.

Results on data sets of BraTS challenge

We also apply our approach on the data sets in the BraTS challenge (<http://www2.imm.dtu.dk/projects/BRATS2012>). The data sets contain separate high-grade (HG) and low-grade (LG) data sets. This results in 4 sets (Real-HG, Real-LG, Synth-HG, Synth-LG).

For convenience of comparison, the process testing our approach is same as that of Zikic et al. [41], which is one of the best performing methods in the challenge. For the real data sets, leave-one-out cross-validation is used: For each test image, the training is performed on all other images from the

data set, excluding the test image itself. For the synthetic images, we perform a leave-five-out cross-validation.

Our segmentation run times is about 30 min per patient (including time for feature extraction). The training of one data set takes approximately one and a half hours on a single desktop PC (including time of two SVM training procedures for tumor and edema, not including time of feature extraction). We summarize the results for the Dice score in Table 1, and the Dice score can be calculated as follows: $Dice = 2 \times Jaccard / (1 + Jaccard)$.

From the Table 4, our method achieves a satisfactory result. With respect to live performance results, our scores are slightly under-performing although not significantly.

Conclusion

Our paper brings together two recent trends in the brain tumor segmentation literature: (1) model-aware similarity and affinity calculations with (2) CRF models with SVM-based evidence terms. In doing so, we make three main contributions. We use superpixel-based appearance models to reduce computational cost, improve spatial smoothness, and solve the

Table 4 Evaluation summary

Method	Dice score	High-grade (real)		Low-grade (real)		High-grade (synth)		Low-grade (synth)	
		Edema	Tumor	Edema	Tumor	Edema	Tumor	Edema	Tumor
SVM + AM	Mean	0.61	0.62	0.43	0.59	0.58	0.86	0.47	0.61
	SD	0.18	0.12	0.21	0.16	0.30	0.05	0.27	0.25
Zikic et al.	Mean	0.70	0.71	0.44	0.62	0.65	0.90	0.55	0.71
	SD	0.09	0.24	0.18	0.27	0.27	0.05	0.23	0.20

data sampling problem for training SVM classifiers on brain tumor segmentation. Also, we develop an affinity model that penalizes spatial discontinuity based on model-level constraints learned from the training data. Finally, our structural denoising based on the symmetry axis and continuity characteristics is shown to remove the false positive regions effectively. Our full system has been thoroughly evaluated on a challenging 20-case GBM and the BraTS challenge data set and shown to systematically perform on par with the state of the art. The combination of the two tracts of ideas yields better performance, on average, than either alone. In the future, we plan to explore alternative feature and classifier methods, such as classification forests to improve overall performance.

Acknowledgments This work was partially supported by the Chinese National Science Foundation (61273241) and the NSF CAREER grant IIS-0845282.

Conflict of Interest The authors have declared that no conflict of interest exists.

References

- Liu J, Udupa JK, Odhner D, Hackney D, Moonis G (2005) A system for brain tumor volume estimation via MR imaging and fuzzy connectedness. *Comput Med Imaging Graphics* 29(1):21–34
- Sled JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 17(1):87–97
- Belaroussi B, Milles J, Carne S, Zhu YM, Benoit-Cattin H (2006) Intensity non-uniformity correction in MRI: existing methods and their validation. *Med Image Anal* 10(2):234
- Madabhushi A, Udupa JK (2005) Interplay between intensity standardization and inhomogeneity correction in MR image processing. *IEEE Trans Med Imaging* 24(5):561–576
- Patel MR, Tse V (2004) Diagnosis and staging of brain tumors. *Semin Roentgenol* 39(3):347–360
- Prastawa M, Bullitt E, Ho S, Gerig G (2004) A brain tumor segmentation framework based on outlier detection. *Med Image Anal* 8(3):275–283
- Phillips W, Velthuizen R, Phuphanich S, Hall L, Clarke L, Silbiger M (1995) Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme. *Magn Reson Imaging* 13(2):277–290
- Clark MC, Hall LO, Goldgof DB, Velthuizen R, Murtagh FR, Silbiger MS (1998) Automatic tumor segmentation using knowledge-based techniques. *IEEE Trans Med Imaging* 17(2):187–201
- Fletcher-Heath LM, Hall LO, Goldgof DB, Murtagh FR (2001) Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artif Intell Med* 21(1–3):43–63
- Warfield SK, Kaus M, Jolesz FA, Kikinis R (2000) Adaptive, template moderated, spatially varying statistical classification. *Med Image Anal* 4(1):43–55
- Kaus MR, Warfield SK, Nabavi A, Black PM, Jolesz FA, Kikinis R (2001) Automated segmentation of MR images of Brain Tumors I. *Radiology* 218(2):586–591
- Prastawa M, Bullitt E, Moon N, Van Leemput K, Gerig G (2003) Automatic brain tumor segmentation by subject specific modification of atlas priors. *Acad Radiol* 10(12):1341–1348
- Wells W III, Grimson WEL, Kikinis R, Jolesz FA (1996) Adaptive segmentation of MRI data. *IEEE Trans Med Imaging* 15(4):429–442
- Guillemaud R, Brady M (1997) Estimating the bias field of MR images. *IEEE Trans Med Imaging* 16(3):238–251
- Corso JJ, Sharon E, Dube S, El-Saden S, Sinha U, Yuille A (2008) Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Trans Med Imaging* 27(5):629–640
- Zhou J, Chan K, Chong V, Krishnan S (2006) Extraction of brain tumor from MR images using one-class support vector machine. In: 27th annual international conference of the engineering in medicine and biology society, 2005. *IEEE-EMBS 2005*. pp 6411–6414
- Xuan X, Liao Q (2008) Automated MRI brain tumor segmentation based on feature extraction. *Comput Eng* 9:070
- Corso J, Yuille A, Sicotte N, Toga A (2007) Detection and segmentation of pathological structures by the extended graph-shifts algorithm. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. pp 985–993
- Schapire RE, Freund Y, Bartlett P, Lee WS (1998) Boosting the margin: a new explanation for the effectiveness of voting methods. *Ann Stat* 26(5):1651–1686
- Dettling M (2004) BagBoosting for tumor classification with gene expression data. *Bioinformatics* 20(18):3583–3593
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Zikic D, Glocker B, Konukoglu E, Criminisi A, Demiralp C, Shotton J, Thomas O, Das T, Jena R, Price S (2012) Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Lecture notes in computer science, vol 7512. pp 369–376
- Liu H, Motoda H, Yu L (2004) A selective sampling approach to active feature selection. *Artif Intell* 159(1):49–74
- Fernández A, García S, Herrera F (2011) Addressing the classification with imbalanced data: open problems and new challenges on class distribution. In: *Hybrid artificial intelligent systems*. Lecture notes in computer science, vol 6678. pp 1–10
- Li SZ (1995) *Markov random field modeling in computer vision*. Springer, New York
- Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20(1):45–57
- Lee CH, Schmidt M, Murtha A, Bistriz A, Sander J, Greiner R (2005) Segmenting brain tumors with conditional random fields

- and support vector machines. In: Computer vision for biomedical, image applications. Lecture notes in computer science, vol 3765. pp 469–478
28. Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. *Int J Comput Vis* 59(2):167–181
 29. Xu C, Corso JJ (2002) Evaluation of super-voxel methods for early video processing. In: IEEE conference on computer vision and pattern recognition (CVPR) 2012, pp 1202–1209
 30. Bauer S, Nolte L-P, Reyes M (2011) Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011. Springer, pp 354–361
 31. Platt J (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv Large Margin Class* 10(3):61–74
 32. Ren X, Malik J (2003) Learning a classification model for segmentation. In: Proceedings of the ninth IEEE international conference on computer vision, 2003. pp 10–17
 33. Manjunath BS, Ma WY (1996) Texture features for browsing and retrieval of image data. *IEEE Trans Pattern Anal Mach Intell* 18(8):837–842
 34. Lee TS (1996) Image representation using 2D Gabor wavelets. *IEEE Trans Pattern Anal Mach Intell* 18(10):959–971
 35. Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. *IEEE Trans Pattern Anal Mach Intell* 23(11):1222–1239
 36. Boykov Y, Kolmogorov V (2004) An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans Pattern Anal Mach Intell* 26(9):1124–1137
 37. Kolmogorov V, Zabih R (2004) What energy functions can be minimized via graph cuts? *IEEE Trans Pattern Anal Mach Intell* 26(2):147–159
 38. Zabih RD, Veksler O, Boykov Y (2004) System and method for fast approximate energy minimization via graph cuts. U. S. Patent 6,744, 923 [p]. 1 June 2004
 39. Delong A, Osokin A, Isack HN, Boykov Y (2012) Fast approximate energy minimization with label costs. *Int J Comput Vis* 96(1):1–27
 40. Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, Henkelman RM, Cusimano MD, Dirks PB (2004) Identification of human brain tumour initiating cells. *Nature* 432(7015):396–401
 41. Zikic D, Glocker B, Konukoglu E, Shotton J, Criminisi A, Ye DH, Demiralp C, Thomas OM, Das T, Jena R, Price SJ (2012) Context-sensitive classification forests for segmentation of brain tumor tissues. In: Proceedings of MICCAI-BRATS (2012)