

An Approach Patch-Based for the Segmentation of Pathologies Application to Glioma Labelling

Syeda Meraj Bilfaqih, Sabahat Khatoon, Mariyam Ayesha Bivi

Department Of Computer Science, King Khalid University, Saudi Arabia

ABSTRACT

In this paper, we describe a novel and generic approach to address fully-automatic segmentation of brain tu-mors by using patch-based voting techniques. In addition to avoiding the local search window assumption, the conventional patch-based framework is enhanced through several simple procedures: an improvement of the training dataset in terms of both label purity and intensity statistics, augmented features to implicitly guide the nearest-neighbor-search, multi-scale patches, invariance to cube isometries, stratification of the votes with respect to cases and labels. A probabilistic model auto-matically delineates regions of interest enclosing high-probability tumor volumes, which allows the algorithm to achieve highly competitive running time despite minimal processing power and resources. This method was evaluated on Multimodal Brain Tumor Image Segmentation challenge datasets. State-of-the-art results are achieved, with a limited learning stage thus restricting the risk of overfit. Moreover, segmentation smoothness does not involve any post-processing.

Keywords : Patch-Based, Glioma, Segmentation.

I. INTRODUCTION

A. Motivation

LIOBLASTOMA is the most severe case of brain Gtumors. Clinical guidelines such as RECIST [1] or RANO [2] are limited to 1D or 2D analysis (maximal diameter and possibly second diameter) of the lesions. However, from tumor growth monitoring to radiotherapy planning, 3D anal-ysis is crucial in the clinical pipeline [3], [4]. Glioblastoma segmentation consists in a 3D delineation of the pathological compartments [5] shown in Figure 1. Manual segmentation is usually complex, subjective and time-consuming. First, glioblastoma exhibit high tumor shape variability. Second, the border between compartments can appear fuzzy, which can lead to a debatable segmentation: inter-rater variability of manual segmentations is in the range 74-85% (Dice over-lap) [5]. Third, the segmentation task requires the simultaneous screening of 3D images acquired with multiple Magnetic Res-onance (MR) sequences (Figure 1). This explains the ongoing interest for automatic

segmentation algorithms, notably within the Multimodal Brain Tumor Image Segmentation (BraTS) benchmark challenge [5].

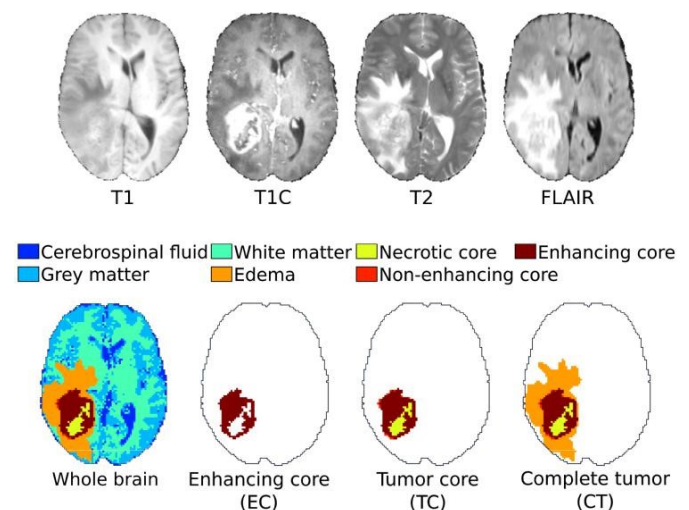


Figure 1. MR channels (top row) ; whole brain segmentation and mutually-inclusive pathological regions (bottom row).

II. METHODS AND MATERIAL

Related Work

1) Glioma segmentation: Most of the automatic glioma segmentation approaches learn offline a discriminative model [5]–[7]: image intensity features are computed, then a machine learning algorithm is trained offline. Most computation time is spent during the learning stage, which should be run again if newly acquired data is annotated. Moreover, results are highly dependent on the choice of features [7], and feature extraction has to be performed at test time. For instance, [7] introduce a cascade of random forest classifiers based on a set of intensity, geometry and asymmetry features. The segmentation is refined using Markov random field regularization. Feature extraction takes about 90 minutes (single-thread implementation) [5], mostly due to the computation of the most relevant features (asymmetry) [7].

The generative approach builds a probabilistic model of observed image intensity given the tissue type. The latent variable is the spatial distribution of healthy tissues and tumor compartments. Prior knowledge includes the location and spatial extent of healthy tissues in an atlas. [8] introduce a generative model for channel-specific pathology segmentation: the appearance of healthy tissues is modelled as a multivariate distribution for all channels simultaneously, while tumor appearance is channel-specific. Tumor location is a latent variable shared between MR channels, which results in tumor segmentations consistent across channels. Since tumor compartments are not explicitly modelled in [8], a discriminative classifier is learnt to transform channel-specific abnormality probability maps into tumor compartment delineations [5]. To reach state-of-the-art results, a strong post-processing is required to deal with the high number of false positives. Running time is 20 minutes [5].

Recently, Kwon [9] achieved top rank among the BraTS benchmark competitors with a semi-automatic framework. First, an expert inputs multiple seed points and radii, and samples one point from each tissue class. A tumor shape prior is generated from the seed points via random walk. A pathological atlas is synthesized by growing tumors from the seed points and combining the result with a normal atlas. Average running time is less

than 10 minutes for the user inputs, plus 85 minutes for the remaining automatic pipeline on a machine with 8 cores.

2) Multi-atlas segmentation: When applied to glioma segmentation, machine learning techniques are confronted with two major problems. First, the amount of training data is usually small: for instance, there are only 20 high-grade training cases for the 2013 BraTS benchmark [5]. Second, most algorithms require a computationally intensive offline learning stage, which can be subject to overfit. Multi-atlas segmentation methods are appealing as they can cope with a small training dataset, and are performed online, which allows a seamless integration of new cases into the training dataset. Atlas-guided segmentation consists in encoding the relationship between image intensities and labels through the use of an atlas, an expert-annotated image [10]. The segmentation of a test case relies on label propagation, the deformable registration of the atlas to the test case. Recently, multi-atlas segmentation [11], which uses several atlases to best capture anatomical variability, has proven successful for the segmentation of healthy brain structures [10], [12]–[14]. However, limitations include i) the high computational cost of non-linear registration, and ii) the assumption that a one-to-one mapping exists between atlas and patient. The development of patch-based segmentation alleviates these two limitations by performing only affine registration and by introducing the concept of local search window to take into account registration errors: since brain structures should be located around the same position after registration, it is sufficient to compare a test patch to annotated patches positioned in a spatial neighbourhood. Patch-based methods have achieved state-of-the-art results for the segmentation of brain structures [15]–[17]. [18] adopt a multi-atlas approach for the detection of cancerous brain regions, based on atlases of healthy brains. This approach relies on the core assumption that there is an equivalence between image outliers and pathological regions. Similar works include outlier detection using a patch-based metric [19]. However, local search cannot be used for the multi-label segmentation of pathologies: tumor patches may not be located in the same region of the human body.

In this paper, we address the automatic segmentation of brain tumors by using multi-atlas patch-based techniques, without any assumption of local search, so

that i) minimal learning is required, which decreases the risk of overfit, especially on a small training dataset, ii) minimal post-processing is required, since segmentation is robust, and spatially consistent labelling is guaranteed by patch overlap, iii) competitive running times are achieved using a single core, and the algorithm could benefit from a computer grid as it is embarrassingly parallel. To obtain a reliable patch-based method which does not rely on the concept of local search window, we combine several methodological advances: feature augmentation based on spatial prior and robust intensity likelihood, selection of training dataset based on label purity and intensity statistics, and a stratification of the votes with respect to cases and labels. In the following, an enhanced patch-based segmentation framework is introduced to accommodate with the presence of a pathology (Section II). Then the algorithm is evaluated on publicly available MR images, and its performance is benchmarked against state-of-the-art methods (Section III). Finally, results and perspectives are discussed (Section IV).

Patch-Based Glioma Segmentation

Conventional patch-based segmentation without any search window constraint is illustrated in Figure 2, along with our contributions. The fundamental assumption is that central voxels of similar patches should have similar labels. Consequently, the label of a test patch could be inferred by finding similar patches in a subset of the training cases and performing a weighted vote: each patch votes for its label with a weight depending on its similarity to the target patch. A probabilistic model is first described to automatically define a region-of-interest (ROI) within the target image. Then three improvements to the conventional framework are proposed: feature augmentation and invariance, training patch selection, and stratification of votes.

A. Notations

Patch-based segmentation relies on a set of training cases $f_{In}; L_{ngn}$ where I_n denotes multi-channel MRI and L_n is a label map. A label l is assigned to a test patch $S(J; x)$ based on comparisons with a database of annotated patches $f_{S(I_n; y)gn; y}$, where y indexes every spatial position in the reference space. Conventionally, a mono-channel patch $S(I; x)$ is a cube of edge length $2r+1$, centered at spatial position $x \in R^3$, consisting of

image intensity values taken from a mono-channel MR image I . The distance $d(S(I; x); S(J; y))$ between 3D mono-channel patches of identical edge length is the canonical L2-norm on $R^{(2r+1)^3}$. A multi-channel patch is the concatenation of mono-channel patches. The squared distance between multi-channel patches is the sum of the squared distances between mono-channel patches.

B. Patch Selection

A probabilistic model for glioma detection 1) Motivation: segmentation speed-up: The automatic definition of ROI, enclosing high-probability tumor regions, allows patch-based segmentation to achieve competitive running time, thanks to patch selection ahead of time. Indeed, patches outside of the target ROI in the target case are directly discarded, which cuts down on time for both patch extraction and patch match. Moreover, patches outside of the training ROI in the training case are similarly discarded since they are unlikely to be relevant matches. In practice, target patch selection has close to no influence on segmentation results: discarded target patches are unlikely to trigger any alarm during segmentation, since the features used for detection (average intensity over a patch) are less specific than the patches used for segmentation.

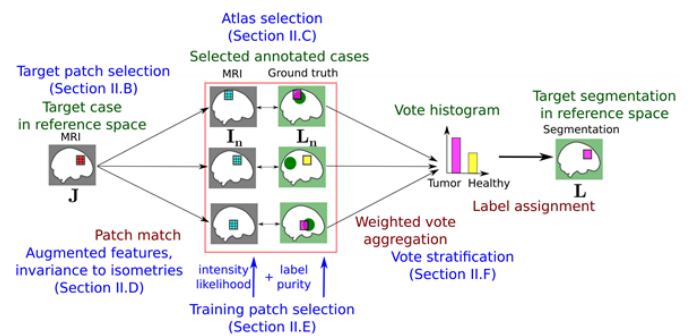


Figure 2. Patch-based glioma segmentation. Our contributions are mentioned in blue. Green disks mark known tumor locations.

2) Definition of a probabilistic model: The idea is to perform a robust clustering of patches into 7 clusters based on average patch intensity. Clusters include cerebrospinal fluid (CSF), grey matter (GM), white matter (WM), necrotic core (Nec.), edema, non-enhancing tumor core (NETC), and enhancing core (EC). Robustness to outliers arising from imaging artefacts, bias field or ground truth mislabelling, is achieved through the use of multivariate Student distributions to model the heavy-tailed distribution of

multi-channel intensity. The proposed model is a mixture of Student distributions [20], with fixed and spatially-varying mixing coefficients, similar to [21]. Each Student distribution is parametrized by $\mu = (\mu; \Sigma; \nu)$, where μ is the mean, Σ the covariance matrix, and ν the number of degrees of freedom. The joint probability is:

$$p(J; L; X) = p(J | L) p(L | X) p(X)$$

where X indexes the coordinates of the brain in the MNI space. The label map L is assumed to be sampled independently for each position x from a generalized Bernoulli distribution with parameters $f(x; \mu)$, our prior regarding the spatial extent of tissue classes. Given a class $L(x) = l$, the multimodal image intensity $J(x) \in \mathbb{R}^4$ is sampled independently from a Student distribution with parameters μ .

3) Estimation of the spatial probabilistic prior: To build atlases of label spatial distribution, every training case is affinely registered to a reference space [22], then, for each label, an atlas of spatial distribution is defined as the voxel-wise average of warped label maps. Atlases are finally symmetrized with respect to the mid-sagittal plane and smoothed by a Gaussian convolution (see examples in Figure 3). For this task, 314 ground truth segmentations from BraTS are used, and the standard deviation of the smoothing Gaussian kernel is 10 mm.

4) Estimation of the parameters of Student distributions:

Multi-channel patch information is summarized by the average intensity of mono-channel patches. Multivariate Student

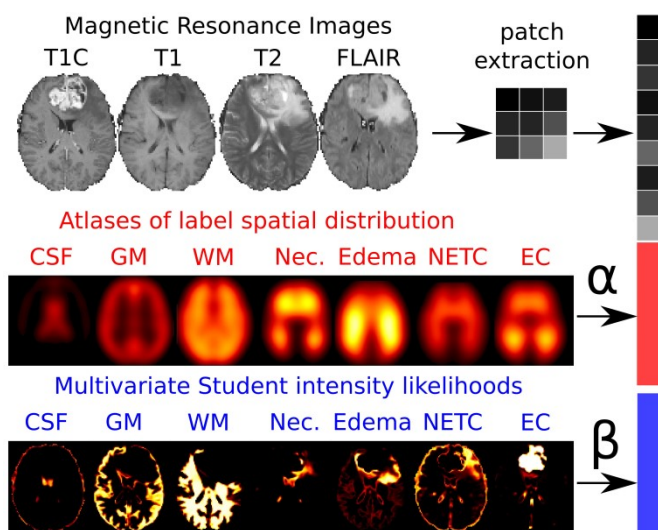


Figure 3 : Feature augmentation. A feature vector is composed of i) intensity patches extracted from 4 MR channels, ii) scalar values from 7 atlases of label spatial distribution (weight), and iii) scalar values from 7 class conditional Student intensity likelihoods (weight).

distributions are fitted using Expectation-Maximization [23] on $6 \times 6 \times 6$ mm³ pure patches, i.e. for which all voxels belong to the same class. Given the optimal parameters μ , class conditional Student intensity likelihood can be computed in closed form for any patch (see Figures 3 and 4).

5) Estimation of the posterior probabilities: The posterior probability of observing label L is given by Bayes' rule:

$$p(L | J; X) / p(J | L) p(L | X)$$

6) Automatic glioma detection: To cope with the lack of specificity of detection features, the maximal posterior map is post-processed by thresholding signed Euclidean distance maps, as shown in Figure 5. The goal is to discard thin

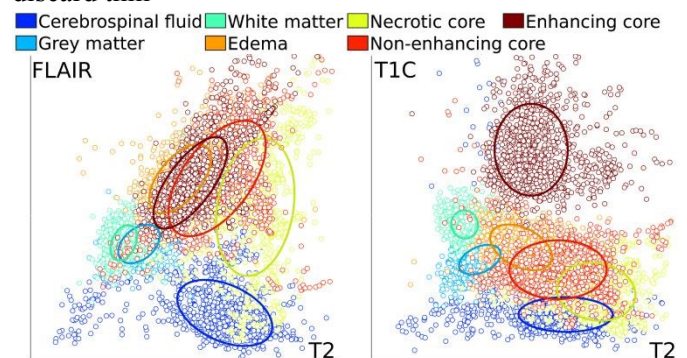


Fig. 4: 2D projections of Student distributions fitted to average intensity of pure patches, with 60% quantiles overlaid.

layers of tumor-detected voxels which are adjacent to healthy-detected tissues, e.g. false positives at the outline of ventricles. Three distance thresholds are used: i) 3 mm from each healthy tissue (CSF, GM, WM), ii) 6 mm from the intersection of the binary masks obtained in step i, and iii) 6 mm from the detected tumor binary mask obtained in step ii.

C. Atlas selection: segmentation scaling to large data sets

Atlas selection, i.e. the pre-selection of a fixed number of training cases similar to the target case, is one solution to scale to large datasets. Higher priority is given to training cases with the same grade as the target. To sort training cases of identical grade, the distance between target case J and training case I_n is the arithmetic mean, for every label l , of the class-specific Hellinger distances between histograms

$(p(L = l | J; X = x))^{1/2}$ and $(p(L = l | I_n; X = x))^{1/2}$. For this purpose, the posterior probabilities detailed in Section II-B5 are used.

D. Feature augmentation for a guided patch match

Patch match consists in finding the most similar patches to a target patch. An augmentation of patch features with additional contextual features can implicitly guide patch match. Three feature augmentations and one feature invariance are proposed.

1) Multi-scale patches: Conventional patches constrain our vision of structures to the scale of the patch. However, gliomas exhibit structures at scales varying from case to case: i) the edema often has the largest extent for high-grade cases, but can be very small compared to the tumor core for low-grade cases ; ii) the distinction between necrotic core and enhancing core may be challenging due to complex borders.

As explained in Figure 6, we have adopted inherently multi-scale patches [24] with two scales: 6 mm (3x3x3 patch) and 18 mm (9x9x9 patch). The central part of the patch is described precisely, while the peripheral part is described by average intensity values over smaller patches, by analogy with the foveal vision. This idea allows to capture longer-range image information since it is computationally efficient: conventional 3x3x3 and multi-scale 9x9x9 patches have a similar memory footprint.

More precisely, conventional mono-channel patches are represented as vectors of the voxel-wise intensity values, e.g. L_n scalar values for an n -D patch of length L . However, for 3D multi-scale mono-channel patches: i) from a given voxel, we consider its 26 neighbouring voxels making a 3x3x3 image patch (this represents 27 scalar values) ; ii) we then consider all the 26 neighbouring conventional 3x3x3 patches and compute

the average intensity for all these neighbouring patches, thus providing 26 additional scalar values. In the end, the feature vector corresponding to a 3D multi-scale and mono-channel patch is composed of 53 scalar intensity values. This process is repeated for each MR channel to process 3D multi-scale and multi-channel patches.

2) Atlases of label spatial distribution: One aspect to study glioma pathogenesis is the analysis of the distribution of gliomas by anatomic locations [25]–[27]. Based on 331 cases among which 47% were glioblastomas, [25] report the majority of gliomas in the frontal and temporal lobes, with 29% and 14% occurrences respectively. Similarly, based on 314 cases among which 82% are glioblastomas, most gliomas in the BraTS benchmark dataset are found in the temporal lobe, as illustrated with the atlas in Figure 7. We augment patch features with the atlases of spatial distribution for each of the 7 labels, in order to implicitly encode patch position in the reference space. The additional feature is a vector in R^7 , which is first normalized so that its sum is equal to the average intensity value reported in Section III-B, then weighted by a coefficient 0. The distance between augmented features is the L2-norm on concatenation of patch values and weighted additional features.

3) Multivariate Student intensity likelihoods: Our suggestion is to augment patch features with robust Student intensity likelihoods, which encode the relative global intensity distance of a patch, with respect to each class taken as a whole. The fit of Student distributions is illustrated in Figure 4. Exactly as in Section II-D2, the additional feature is a vector in R^7 , which is first normalized (unit normalization, followed by the multiplication by a scalar value so that the L1 norm is the average intensity value reported in Section III-B) and then weighted by a coefficient 0.

4) Invariance to cube isometries: The canonical L2 distance is sensitive to rotation or symmetry of the patches. This sensitivity is justified for healthy brain structures since rotated patches can be unrealistic. However, since the brain is roughly symmetric with respect to the mid-sagittal plane, the application of sagittal plane symmetry to healthy training patches results in plausible healthy patches. For pathological training patches, it is desirable that the similarity measure is insensitive to rotation and symmetry: gliomas do not seem to exhibit any general trend in terms of texture anisotropy. More-

over, the training dataset is small, so plausible configuration of pathological patches are missing. Cube isometries consist of certain rotations under which the cube is invariant, plus their composition with central symmetry. The application of the 48 cube isometries to pathological training patches allows to generate additional plausible configurations, which leads to an invariant patch distance.

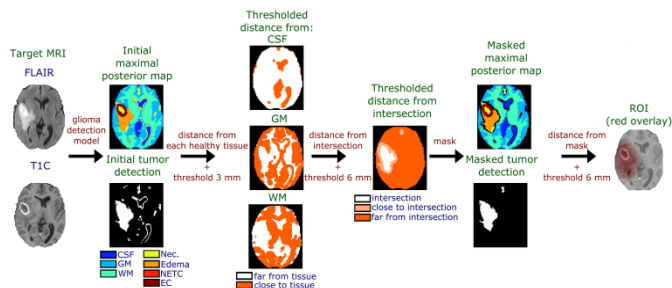


Figure 5: Glioma detection post-processing. Euclidean distances from each healthy tissue are thresholded and intersected to determine the tumor bulk. The ROI is the result of a dilatation of the tumor mask using two additional distance maps.

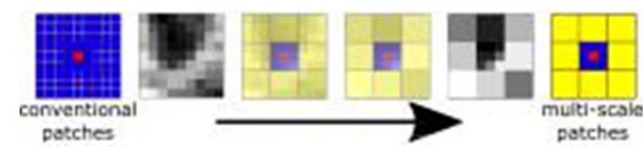


Figure 6: Multi-scale patch as described in [24]. In red: central voxel. In blue: intensity described at the same scale as the central voxel. In yellow: intensity described at a coarser scale, typically average intensity over the coarser area.

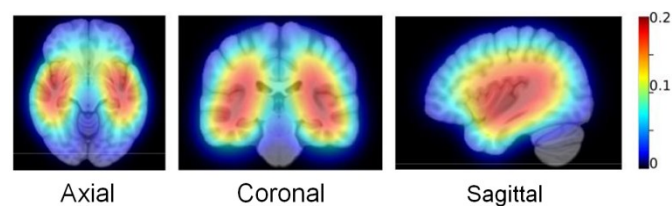


Figure 7 : Atlas of spatial distribution of complete tumor super-imposed on a template. Colormap: percentage of occurrences.

E. Training patch selection for a robust patch match

Manual segmentations are prone to errors due to tumor complex appearance and shape. Training patch selection consists in trimming the training data, so as to increase its robustness to mislabelled patches.

1) Data pruning based on class conditional Student intensity likelihood: The most representative patches lie near their class centroid, so patches whose Student likelihood lie outside of a 60%-quantile could be discarded as dubiously labelled or least representative (Figure 4).

2) Data pruning based on patch label purity: Patch label purity is defined as the percentage of voxels sharing the same label as the central voxel in a patch. The fundamental assumption of patch-based methods may not hold for patches of low purity. Different purity thresholds are tested in Section III-F4.

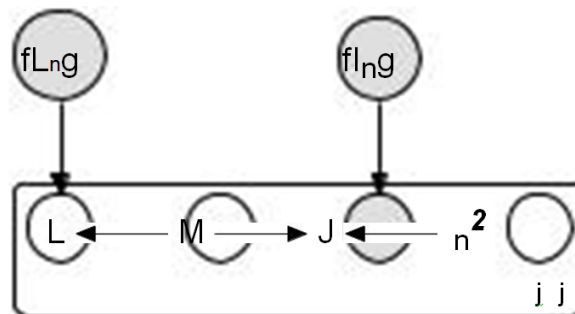


Figure 8. Graphical model for patch-based segmentation. A membership variable $M : \{1, \dots, N_g\}$ is sampled at every position x in the MNI space to encode the training case n to sample from, and the spatial offset y . The observed intensity $J(x)$ results from the sampling of a patch $S(J; x)$ from a Gaussian distribution with $S(I_n; y)$ for mean and isotropic scaling $n^2(x)$ for variance. The label L at position x is the central label $L_n(y)$.

F. Vote stratification for a robust vote aggregation

1) Conventional model for patch-based segmentation:

Conventional patch-based segmentation is a variation of local weighted voting [10], [28] within Bayesian modeling of multi-atlas segmentation [29], as recalled in Figure 8. The target image is the result of sampling patches from atlases $(I_n; L_n)$ at different positions y in the reference space. The membership index M encodes, for every position $x \in \Omega$, both the atlas index n and the position y to sample from: $M(x) = (n; y)$. If M were known, the target segmentation L would be given by labels $L_n(y)$ at central voxels of sampled patches. However, since M is not observed, the inference has to be performed by marginalizing over M [29]. The exact

marginalization consists in a weighted vote involving patches at every position y in every atlas n , with weights proportional to both the probability that the training patch belongs to class $L_n(y)$, and the intensity likelihood of the training patch.

Assuming the membership index $M(x)$ is independent and identically uniformly distributed $p(M) = (N \ j \ j) \ j \ j$, the marginalization over M [29] is given voxel-wise by:

$$L(x) = \underset{l \in \{1, \dots, L\}}{\operatorname{argmax}} \sum_{n=1}^N p_n(L(x) = l \mid L_n; y) \quad (x)$$

where: i) since label maps of training cases are warped from patient space into a reference space, the ground truth can be represented as probabilistic label map $L_n(y)$, or as a binary label map $L_n(y)$ obtained by binarizing $L_n^f(y)$. The probability of belonging to class $L_n(y)$ is $p_n(L(x) = l \mid L_n; y)$, which is either the l -th component of $L_n(y)$ (probabilistic label), or the indicator function $1_{L_n(y)=l}$ (binarized label), and ii) a multivariate Gaussian probability density function

$$p_n(S(J; x) \mid L_n; y) = \frac{1}{\exp(-\frac{1}{2} (S(J; x) - S(L_n; y))^T \Sigma^{-1} (S(J; x) - S(L_n; y)))}$$

with $p_n(x) = \min_y d(S(J; x); S(L_n; y))$

2) Vote stratification for an approximate marginalization:

The exact marginalization over M would require a high number of comparisons between image patches, which would result in prohibitive computation time. Conventional patch-based methods approximate the marginalization step with K -nearest-neighbor patch match or local search window.

For pathology segmentation, we propose to use a stratified K -nearest neighbor approach in order to improve the robustness of the approximate marginalization. More precisely, the set of all patches from all training patients may be partitioned into Q sets of patches which match a number of criteria such as the patient Id , and label type. Given a test patch, the final vote for its label combines the contributions of the K closest patches within each of the Q sets leading to $K \cdot Q$ votes. Vote stratification can be seen as a way to normalize the vote to limit bias in the training dataset. For this application, we chose to stratify the sets according to patient Id and label type such that $Q = N \cdot L$

and $K = 1$. Indeed, this allows to cope with the variability in tumor size between patients, and the unbalanced number of patches between labels. Moreover, this approximation sounds reasonable since, in practice, we verified that taking into account strictly more than $K = 1$ neighbor does not affect segmentation results. In the case of binary label maps, the proposed approximate marginalization consists in:

$$L(x) = \underset{l \in \{1, \dots, L\}}{\operatorname{argmax}} \sum_{n=1}^N p_n(S(J; x) \mid L_n; y_n; \mathbf{1}(l)) \quad (x)$$

where $y_n, l(c)$ is the position in reference space of the closest patch of label c found in the training case $(L_n; L_n)$. An implementation of this stratification scheme is given in Algorithm 1.

In the case of probabilistic label maps, the equation for an approximate marginalization differs slightly due to the fact that

Algorithm 1: Stratified patch-based segmentation. The blue dashed line marks an embarrassingly parallel for-loop.

```

pre-select N training cases similar to J
foreach voxel x do
  evaluate the features S(J; x)
  foreach training case j do
    initialize the vote vector v_xj = 0 2 R^L
    foreach class l 2 f1; ::::; Lg do
      pre-select K closest patches of label l in I_j
      foreach annotated patch k 2 f1; ::::; Kg do
        evaluate the features S(l_j; y_j;k)
        accumulate the vote:
          v_xj(l) = v_xj(l) + exp( - (jS(J;x) - S(l_j;y_j;k))^2 / (2 * sigma^2) )
      foreach class l 2 f1; ::::; Lg do
        evaluate the probability p_xj(l) = v_xj(l) / sum(v_xj(l))
    foreach class l 2 f1; ::::; Lg do
      evaluate the probability p_x(l) = sum_j p_xj(l)
    set the label: L(x) = argmax_l p_x(l)
  
```

$p_n(L(x) = l \mid L_n; y)$ is no longer an indicator function:

$$L(x) = \underset{l \in \{1, \dots, L\}}{\operatorname{argmax}} \sum_{n=1}^N p_n(L(x) = l \mid L_n; y_n; \mathbf{1}(c))$$

In practice, the only impact of considering probabilistic label maps L_n lies in the vote accumulation step in Algorithm 1. The update becomes a for-loop over classes $c \in \{1, \dots, L\}$:

$$v_{x,j}(c) = v_{x,j}(c) + p_j(L(x) = c \mid L_j; y_j; k) \exp(-\frac{1}{2} (S(J; x) - S(l_j; y_j; k))^2 / \sigma^2)$$

III. RESULTS AND DISCUSSION

RESULTS

A. Dataset and Evaluation

Four MR channels commonly acquired for glioma assessment are available in 1 mm isotropic resolution for each case: pre-contrast T1-weighted image (T1), contrast-enhanced T1-weighted (T1C), T2-weighted (T2), and T2-FLAIR MR images. Details about MRI acquisition and processing, manual delineation protocol, and expert consensus can be found in [5].

The proposed glioma segmentation algorithm is evaluated on publicly available MR images of high-grade (HG) and low-grade (LG) gliomas, made available for training and testing in the 2013 and 2014 editions of the MICCAI BraTS bench-mark [5]. Ground truth is only publicly available for 2013 and 2014 Evaluation datasets. The 2014 Evaluation dataset is used for the construction of atlases of label spatial extent and to fit Student distributions. Unless explicitly mentioned otherwise, the training dataset used for patch match in this article is the 2013 Evaluation dataset for a fair comparison to other competitors. For validation on training cases, a comprehensive leave-one-out procedure is applied to exclude the target image from the training dataset at every stage of the method, namely i) for atlas of label spatial extent construction, ii) to fit Student

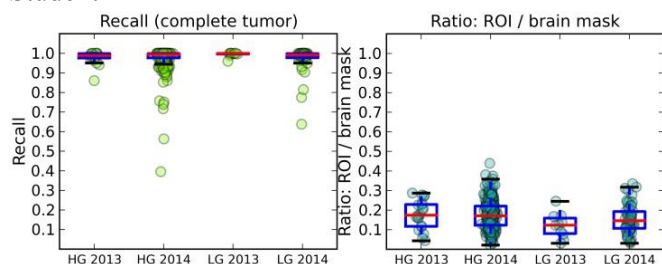


Fig. 9: Left: recall of glioma detection for the complete tumor, reported for cases grouped by year and grade. Right: volume ratio of detected ROI over the brain mask.

distributions, iii) during patch-match. This way, the target case cannot result in any overtraining that would be due to a favourably-biased glioma detection, atlas selection, feature augmentation, or patch match.

B. Pre-processing pipeline

Every image is affinely registered to an MNI atlas and warped to the same reference space. A resampling to 2 mm isotropic resolution is performed to decrease computation time at little accuracy cost: most MR channels were not acquired in 1 mm isotropic resolution, and interpolation artefacts are already visible on few MRI. A robust image normalization ensures that inter-patient patch distance is meaningful: non-zero intensity are clipped below 1% and above 99% quantiles, then mean and standard deviation are set to the values reached by averaging over training data, respectively 360 and 120. Image normalization is performed separately for each MR channel, to allow an equal contribution of each channel in the patch distance. Finally, for the training cases, only the segmentation of the tumor compartments is originally available, so healthy tissues are automatically segmented using FSL FAST [30] to avoid under-segmentation by partitioning the background [31].

C. Glioma detection benchmark

ROI are automatically defined by the glioma detection algorithm described in Section II-B. Recall, i.e. the proportion of tumor voxels which are successfully enclosed in the ROI, is evaluated on the 2013 and 2014 training datasets. Considering all cases, i) regions-of-interest occupy less than 17% of the brain masks, which effectively results in lower computational burden, ii) recall is higher than 97% for every tumor region. Individual results are summarized in Figure 9.

D. Post-processing pipeline for glioma segmentation

Segmentation results are smooth by design due to patch overlap, so post-processing is limited. The pipeline consists in warping back the probabilistic label maps to the 1 mm isotropic patient space, and then keeping at most the two biggest connected components of the segmentation: if the volume of the second component is less than 3 cm³, only the first component is kept.

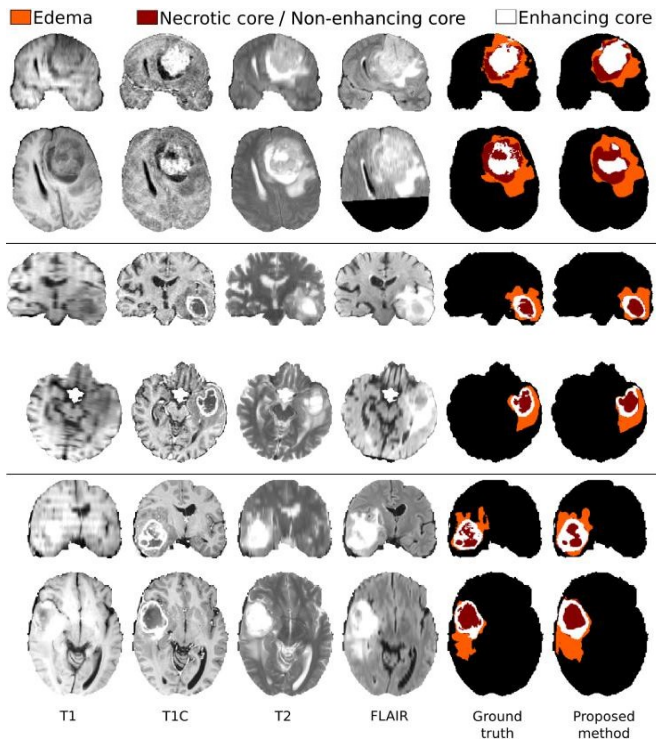


Figure 10: Segmentation of 3 HG cases (coronal and axial views).

E. Segmentation benchmark

The proposed approach is benchmarked on the 2013 Chal-enge dataset against the top-performing documented meth-ods, according to the ranking found on the MICCAI 2013 BraTS benchmark website [5] on April 28, 2015: [9] is the highest ranked semi-automatic method, with Dice over-lap of 88%/83%/72% for CT/TC/EC respectively ; [7] is the highest ranked automatic method, with Dice overlap of 87%/78%/74%. Our automatic method achieves state-of-the-art performance, with Dice overlap of 87%/77%/73%, in a shorter running time. Fast approximate nearest-neighbour search, especially multiple randomized k-d trees for high di-mensional data [32], allows a single-thread implementation of the proposed patch-based segmentation to achieve competitive running times: i) 16 minutes on average for HG, and ii) about 1 minute for LG due to fewer selected atlases, using a single core (2.66 GHz) and less than 6 GB of RAM. Moreover, given access to a computer grid with at least as many cores as there are selected atlases, the proposed stratification scheme allows to achieve running times close to one minute. Segmentation results1 are shown in Figure 10, and box plots2 are displayed for HG in Figure 11.

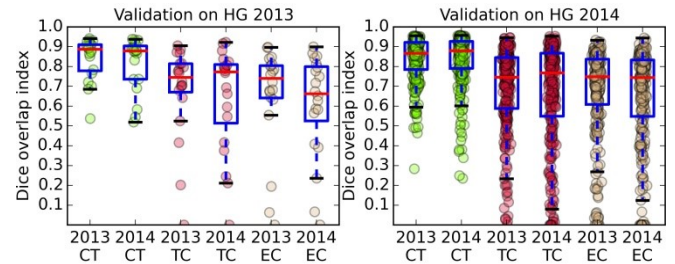
F. Parameter setting

The optimal set of parameters is determined, using leave-one-out on the 2013 training dataset of HG cases, by changing one parameter at a time and defining its optimal value based on

- 1 All the segmentation results are shown in supplementary material for the 2013 validation dataset.
- 2 A table with average assessment measures is given in supplementary material for all the datasets.

The proposed payment system combines the Iris recognition with the visual cryptography by which customer data privacy can be obtained and prevents theft through phishing attack [8]. This method provides best for legitimate user identification. This method can also be implemented in computers using external iris recognition devices.

Dice overlap



Hausdorff distance

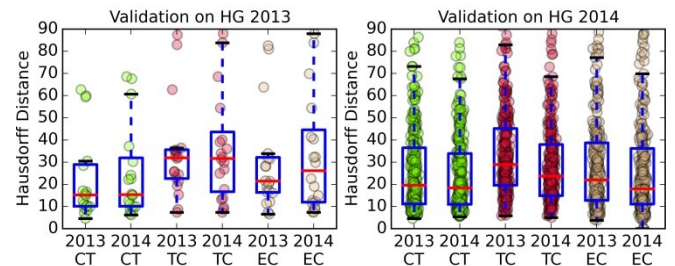


Figure 11 : Box plots of Dice overlap and Hausdorff distance for HG cases. The test dataset consists of the 2013 (left) and 2014 (right) MICCAI BraTS benchmark Evaluation datasets. In abscissa of each plot, 2013/2014 indicates the year used for the training atlases, e.g. 2013 for the 2013 Evaluation dataset ; CT/TC/EC denotes respectively complete tumor, tumor core, and enhancing core.

Dice overlap. No post-processing is applied. Moreover, binary label maps are used, which results in lower Dice overlap than with probabilistic label maps.

1) Atlas selection: number of selected training cases:

Given a target case, a number of similar training cases are selected to perform patch-based segmentation. The selection first considers training cases of the same grade as the target, and once the training dataset is exhausted, training cases of a different grade start to be selected. The influence of atlas selection count, i.e. the number of selected training cases, is studied in Figure 12. For HG target, an optimum of Dice overlap is reached for 28 training cases if we only consider the complete tumor, and for 24 training cases (20 HG, 4 LG) if we simultaneously consider all the tumor regions. For LG target, Dice overlap is optimal once 3 LG training cases are selected, and then tends to decrease, especially as HG cases are selected: it is mostly detrimental to segment LG tumor core based on HG training cases.

2) Feature: patch width, multi-scale patches, and invariance to cube isometries or sagittal plane symmetry: Different patch widths were tested, as shown in Figure 13. In the left pane for which features are not invariant to cube isometries or sagittal plane symmetry, the optimal patch width is 3, which is the smallest tested width. The use of multi-scale patches, including a 3-voxel-wide central patch, results in Dice overlaps similar to these obtained with conventional 5-voxel-wide patches. In the right pane for which features are invariant, the optimal patch width for conventional patches is 5. Comparable yet slightly better results are obtained with multi-scale patches. Invariance decreases Dice overlap for the smallest conventional patches, but increases Dice overlap for larger conventional patches and for multi-scale patches. Intuitively, matching larger patches is sufficiently constrained so that patch match actually benefits from an augmentation of the training dataset via cube isometries or sagittal plane symmetry. In the end, 9-voxel-wide multi-scale patches, with a 3-voxel-wide central patch, are used along invariance to cube isometries for pathological training patches, and invariance to sagittal plane symmetry for healthy training patches.

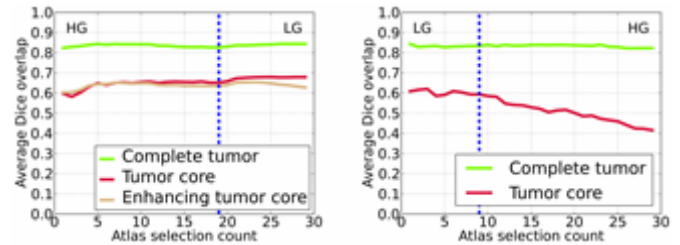


Figure 12. Influence of atlas selection count, in leave-one-out. Vertical lines mark transitions between tumor grades of selected cases.

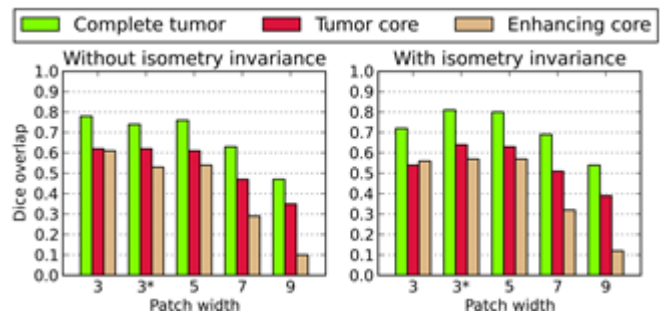


Figure 13: Influence of patch width, multi-scale enhancement, and invariance to cube isometries on the average Dice overlap for HG target cases. The "3*" index corresponds to multi-scale patches with a central patch of indicated width. The right part uses invariance to cube isometries (for pathological patches) and sagittal plane symmetry (for healthy patches).

3) Feature augmentation weights and : With feature augmentation, a weighted L2-norm replaces patch distance, with weights and respectively for atlases of label spatial distribution and for class conditional Student intensity likelihoods. The influence of both weights is studied in Figure 14. Increasing leads to higher Dice overlap, with visually similar segmentation results in the range [8; 14]. When tends to infinity, the segmentation is only driven by label spatial priors, and Dice overlap measures tend to 0. Decreasing leads to higher Dice overlap. Indeed, class conditional Student intensity likelihoods do not allow to discriminate between classes in areas where clusters overlap (Figure 4). When tends to infinity, Dice overlap measures fall off to 42%, 26% and 32% for complete tumor (CT), tumor core (TC), and enhancing core (EC) respectively. In the end, = 12 and = 2 are chosen, since a low non-zero value for seems to result in slightly smoother segmentations.

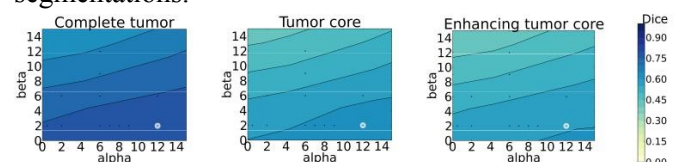


Figure 14: Influence of feature augmentation weights (,) on the average Dice overlap for HG target cases. Chosen parameter values (,) are circled in white.

4) Training patch selection: quantile of Student distributions and minimal patch label purity: The effect of training patch selection is mostly visible if the pre-processing pipeline accumulates approximations (interpolation of images, and bi-narization of warped label maps). In our case, its influence on Dice overlap is marginal for conventional or multi-scale patches. Dice overlap for enhancing core could benefit from a higher minimal patch label purity, but this would decrease Dice overlap for complete tumor and tumor core. In the end, the quantile of Student distributions is equal to 60%, and the minimal patch label purity is equal to 2/3, to ensure shorter running time at the cost of a marginal Dice overlap decrease.

5) Vote aggregation: stratification scheme: The proposed stratification scheme is compared to simpler schemes, with the same number of patch matches for a fair comparison. No atlas selection is performed: the number of training cases is 29 due to leave-one-out. The four schemes are: A) stratification with respect to training cases (29) and labels (7): 1 patch match, B) stratification only with respect to labels (7): 29 patch matches,

C) stratification only with respect to training cases (29): 7 patch matches, D) no stratification: 203 patch matches.

Scheme C is comparable to our naive patch-based method [33], which ranked 5th in 2013, and would rank 18th as of April 2015. The definition of n_2 was different, votes were heuristically scaled, and results required stronger post-processing. On the 2013 training data in leave-one-out, Dice overlap measures were similar to those reported for scheme C: 79%, 60% and 59% for CT, TC and EC respectively.

As shown in Figure 15, a stratification with respect to labels (scheme B) decreases Dice overlap as compared to no stratification (scheme D). On the other hand, a stratification with respect to training cases (scheme C) consistently leads to higher Dice overlap. The situation regarding the tumor core on LG cases is particular since the distinction between edema and tumor core for this grade is especially debatable. Finally, for HG cases, a

stratification with respect to both labels and training cases (scheme A) leads to 1% and 2% higher Dice overlap for CT and TC respectively, and 1% lower Dice overlap for EC. In the end, the chosen scheme is A.

DISCUSSION

A. Limitations of the approach

Overall running time is highly competitive, mostly due to resampling to 2 mm isotropic resolution, target and training patch selection, and atlas selection. However, these methods show some limitations. First, image resampling leads to a loss of the finer structures, which can be detrimental to the segmentation of the enhancing core. For instance, in the second row in Figure 10, the enhancing core signal is subtle in TIC and only partially retrieved by our method. Second, target patch selection depends on a few automatic post-processing steps which could fail and hinder the segmentation process. Third, training patch selection leads to a decrease of Dice overlap, although very marginal in our experiments. Fourth, atlas selection could not be properly studied since the 2013 training dataset is small: the optimal atlas selection count would differ with a larger training dataset.

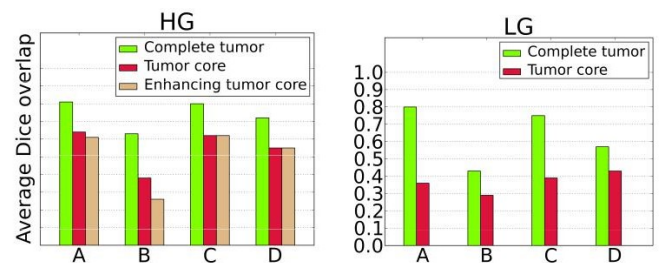


Figure 15. Comparison of stratification schemes A to D.

Patch distance is ultimately the canonical L2-norm. Based on the mean and covariance of the Student distributions, Mahalanobis patch distances have been tested but resulted in a radical drop of performance. This might be due to an inconsistent definition of the minimal patch distance $n_2(x)$ in the case of class-specific patch distances.

Patch matching heavily relies on fast approximate nearest neighbor search for high-dimensional data. In our experiments, approximate search resulted in segmentation performance indistinguishable from exact nearest neighbor search. However, this might be due to the fact that our approach only considers distances between patches to perform the segmentation. This

could lead to drastically different results for any application which would make use of the nearest-neighbor per-se.

B. Limited risk of overfit

Our approach is based on patches extracted from the 2013 training dataset, for which the ground truth is obtained as a consensus of four manual expert segmentations. The 2014 training dataset is only used for validation purpose, and for a few learning stages: the creation of atlases of label spatial distribution, and the fit of Student distributions to average patch intensity. As shown in Figure 11, we tried to base our approach on patches extracted from the 2014 training dataset, but segmentation results were less satisfactory: average Dice overlap was slightly lower for 2013 target cases, and slightly higher for 2014, however Dice overlap variance was higher for both years. One possible explanation is that the ground truth of the 2014 training dataset, which was obtained as a consensus of randomly-selected top-ranking algorithms, could be less reliable. Another explanation is that the two types of delineations (BraTS 2013 and 2014) are based on slightly different assumptions. Nevertheless, segmentation results obtained on the 2 Challenge datasets are consistent, and assessment measures are similar, irrespective of the choice of the training dataset (Figure 11), which seems to indicate that the proposed method does not suffer from overfitting.

Finally, the proposed approach has lead to state-of-the-art brain tumor segmentation results on the BraTS benchmark, and appears to work fine on the clinical data of our collaborators. However, a larger test dataset, with a reliable and hidden expert ground truth, would be necessary to be able to distinguish between the performances of the top-ranked algorithms of the BraTS benchmark. More importantly, given the amount of machine learning segmentation methods, it is essential that test datasets are made available through third-parties as is the case with the BraTS benchmark: it is not uncommon that learning-based procedures overfit, which makes irrelevant any comparison of segmentation results on the training datasets.

IV. SUMMARY AND CONCLUSIONS

We presented a generic approach to adapt patch-based techniques to the segmentation of pathological cases.

This approach has lead to state-of-the-art brain tumor segmentation results in a fully automatic setting. The procedure requires very limited to no prior learning, which limits the risk of overfit and can take advantage of a constantly growing database of annotated cases. Due to patch overlap, segmentation results only require minimal post-processing, which confirms the robustness of the proposed segmentation approach. Patch extraction and matching benefits from automatic patch selection ahead of time, and fast approximate nearest-neighbor search, which results in a highly competitive overall running time. With stratification, patch match is effectively embarrassingly parallel, which results in even shorter running times (close to one minute) using a computer grid.

The possibility to automatically define precise regions of interest in a matter of seconds, using a probabilistic model similar to the one proposed for glioma detection, with a spatially-varying prior learnt on the training data, could be of interest to other medical imaging algorithms. The proposed approach could also be incorporated within the Modality Propagation framework [34], therefore opening new perspectives for the application of patch-based methods to the simulation of medical images including brain pathologies.

V. ACKNOWLEDGEMENTS

Part of this work was funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

VI. REFERENCES

- [1]. E. Eisenhauer, P. Therasse, J. Bogaerts, L. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney et al., "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)," *European Journal of Cancer*, vol. 45, no. 2, pp. 228–247, 2009.
- [2]. P. Y. Wen, D. R. Macdonald, D. A. Reardon, T. F. Cloughesy, A. G. Sorensen, E. Galanis, J. DeGroot, W. Wick, M. R. Gilbert, A. B. Lassman et al., "Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group," *Journal of*

- Clinical Oncology, vol. 28, no. 11, pp. 1963–1972, 2010.
- [3]. E. D. Angelini, O. Clatz, E. Mandonnet, E. Konukoglu, L. Capelle, and H. Duffau, "Glioma dynamics and computational models: a review of segmentation, registration, and in silico growth algorithms and their clinical applications," *Current Medical Imaging Reviews*, vol. 3, no. 4, pp. 262–276, 2007.
- [4]. E. Mandonnet, S. Wait, L. Choi, and C. Teo, "The importance of measuring the velocity of diameter expansion on MRI in upfront management of suspected WHO grade II glioma—Case report," *Neurochirurgie*, vol. 59, no. 2, pp. 89–92, 2013.
- [5]. B. Menze, M. Reyes, K. Van Leemput et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BraTS)," *Medical Imaging, IEEE Transactions on*, vol. PP, no. 99, pp. 1–33, December 2014.
- [6]. D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shot-ton, O. Thomas, T. Das, R. Jena, and S. Price, "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," in *MICCAI 2012*. Springer, 2012, pp. 369–376.
- [7]. N. J. Tustison, K. Shrinidhi, M. Wintermark, C. R. Durst, B. M. Kandel, J. C. Gee, M. C. Grossman, and B. B. Avants, "Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR," *Neuroinformatics*, pp. 1–17, 2014.
- [8]. B. H. Menze, K. Van Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland, "A generative model for brain tumor segmentation in multi-modal images," in *MICCAI 2010*. Springer, 2010, pp. 151–159.
- [9]. D. Kwon, R. T. Shinohara, H. Akbari, and C. Davatzikos, "Combining generative models for multifocal glioma segmentation and registration," in *MICCAI 2014*. Springer, 2014, pp. 763–770.
- [10]. J. E. Iglesias and M. R. Sabuncu, "Multi-Atlas Segmentation of Biomedical Images: A Survey," *Medical Image Analysis*, 2015. Online]. Available: <http://arxiv.org/abs/1412.3421>
- [11]. T. Rohlfing, D. B. Russakoff, and C. R. Maurer Jr, "Extraction and application of expert priors to combine multiple segmentations of human brain tissue," in *MICCAI 2003*. Springer, 2003, pp. 578–585.
- [12]. R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.
- [13]. P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, 2009.
- [14]. M. J. Cardoso, K. Leung, M. Modat, S. Keihaninejad, D. Cash, J. Barnes, N. C. Fox, S. Ourselin, A. D. N. Initiative et al., "STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation," *Medical image analysis*, vol. 17, no. 6, pp. 671–684, 2013.
- [15]. F. Rousseau, P. A. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 10, pp. 1852–1862, 2011.
- [16]. P. Coupe, J. V. Manjon, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.
- [17]. J. E. Romero, J. V. Manjon, J. Tohka, P. Coupe, and M. Robles, "NABS: Non-local Automatic Brain Hemisphere Segmentation," *Magnetic resonance imaging*, 2015.
- [18]. A. J. Asman, B. Landman et al., "Out-of-atlas labeling: A multi-atlas approach to cancer segmentation," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE, 2012, pp. 1236–1239.
- [19]. M. J. Cardoso, C. H. Sudre, M. Modat, and S. Ourselin, "Template-Based Multimodal Joint Generative Model of Brain Data," in *Information Processing in Medical Imaging*. Springer International Publishing, 2015, pp. 17–29.
- [20]. M. Svensen and C. M. Bishop, "Robust Bayesian mixture modelling," *Neurocomputing*, vol. 64, pp. 235–252, 2005.
- [21]. N. Shiee, P.-L. Bazin, J. L. Cuzzocreo, A. Blitz, and D. L. Pham, "Segmentation of brain images using adaptive atlases with application to ventriculomegaly," in *Information Processing in Medical Imaging*. Springer, 2011, pp. 1–12.
- [22]. J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G.

- Simpson, B. Pike et al., "A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 356, no. 1412, pp. 1293–1322, 2001.
- [23]. C. Liu and D. B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19–39, 1995.
- [24]. C. Wachinger, M. Brennan, G. C. Sharp, and P. Golland, "On the Importance of Location and Features for the Patch-Based Segmentation of Parotid Glands," in *MICCAI Workshop on Image-Guided Adaptive Radiation Therapy*, 2014. [Online]. Available: <http://hdl.handle.net/10380/3472>
- [25]. S. Larjavaara, R. Mantyla, T. Salminen, H. Haapasalo, J. Raitanen, J. Jaaskelainen, and A. Auvinen, "Incidence of gliomas by anatomic location," *Neuro-oncology*, vol. 9, no. 3, pp. 319–325, 2007.
- [26]. H. Duffau and L. Capelle, "Preferential brain locations of low-grade gliomas," *Cancer*, vol. 100, no. 12, pp. 2622–2626, 2004.
- [27]. S. Parisot, H. Duffau, S. Chemouny, and N. Paragios, "Graph based spatial position mapping of low-grade gliomas," in *MICCAI 2011*. Springer, 2011, pp. 508–515.
- [28]. W. Bai, W. Shi, D. P. O'Regan, T. Tong, H. Wang, S. Jamil-Copley, N. S. Peters, and D. Rueckert, "A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 7, pp. 1302–1315, 2013.
- [29]. M. R. Sabuncu, B. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 10, pp. 1714–1729, 2010.
- [30]. Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 1, pp. 45–57, 2001.
- [31]. C. Wachinger and P. Golland, "Atlas-Based Under-Segmentation," in *MICCAI 2014*. Springer, 2014, pp. 315–322.
- [32]. M. Muja and D. G. Lowe, "Scalable Nearest Neighbor Algorithms for High Dimensional Data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, 2014.
- [33]. N. Cordier, B. Menze, H. Delingette, and N. Ayache, "Patch-based Segmentation of Brain Tissues," in *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, 2013, pp. 6–17.
- [34]. D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, and E. Konukoglu, "Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization," in *MICCAI 2013*. Springer, 2013, pp. 606–613.