UNCERTAINTY QUANTIFICATION IN BRAIN TUMOR SEGMENTATION USING CRFs AND RANDOM PERTURBATION MODELS

Esther Alberts[1,2*], *Markus Rempfler*[1*], *Georgina Alber*[2], *Thomas Huber*[2], *Jan Kirschke*[2], *Claus Zimmer*[2] *and Bjoern H. Menze*[1 †]

[1] Department of Computer Science, Technische Universität München, Munich, Germany

[2] Neuroradiology, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany

ABSTRACT

Medical image segmentation is a challenging task and algorithms often struggle with the high variability of inhomogeneous clinical data, demanding different parameter settings or resulting in weak segmentation accuracy across different inputs. Assessing the uncertainty in the resulting segmentation therefore becomes crucial for both communicating with the end-user and calculating further metrics of interest based on it, for example, in tumor volumetry.

In this paper, we quantify segmentation uncertainties in a energy minimisation method where computing probabilistic segmentations is non-trivial. We follow recently proposed work on random perturbation models that enables us to sample segmentations efficiently by repeatedly perturbing the energy function of the conditional random field (CRF) followed by maximum a posteriori (MAP) inference. We conduct experiments on brain tumor segmentation, with both voxel and supervoxel perturbations, and demonstrate the benefits of probabilistic segmentations by means of precision-recall curves and uncertainties in tumor volumetry along time.

Index Terms— Random MAP perturbations, conditional random fields, uncertainty quantification, medical image segmentation

1. INTRODUCTION

Medical image segmentation is a well-studied field involving the delineation of cells, tissues, organs and pathological structures. The major difficulties in this field are the inherent ill-posed nature of the segmentation task as well as the high variability of the data observed across subjects (due to tissue heterogeneities), across data acquisitions (due to imaging artifacts) and across medical sites (due to different scanning protocols). Therefore, segmentation algorithms often have to make decisions in the presence of uncertainty. However, validation usually only relies on the hard segmentations generated by the algorithm, and little prior work has been done to estimate and report the uncertainty of the results. In this paper, we want to emphasize that uncertainty in medical image segmentation is a valuable concept and can play a critical role in the evaluation and validation of an algorithm, as well as in clinical practice, where uncertainty can provide useful information in several applications such as disease diagnosis, therapy guidance and radiotherapy planning.

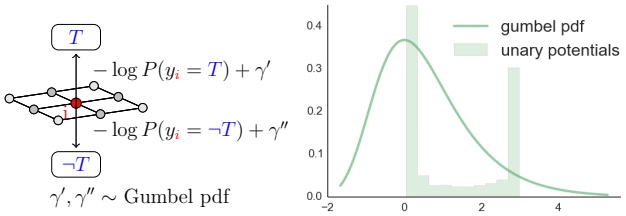

Fig. 1. *Left:* Gumbel samples perturb the unary potentials in between graph nodes i and label terminals T (tumor) and $\neg T$ (non-tumor). *Right:* range of Gumbel samples compared to the non-perturbed unary potentials.

The conventional method to assess segmentation uncertainty is by using probabilistic segmentation models. In the past years, graphical models have become very popular, as they allow to incorporate spatial, temporal and inter-image coherence. However, they make it difficult to go beyond hard segmentation. Traditionally, Markov chain Monte Carlo (MCMC) methods are used for sampling purposes. These methods are computationally expensive, suffer from burn-in periods and lack scalability. Therefore, sampling directly over the entire voxel grid is to be avoided, and instead, one would rather sample from parametric representations as in [1] or use blockwise updates as in [2]. Alternatively, [3] proposed to use min-marginal energies to quantify voxel-specific uncertainties, but it becomes expensive for volumetric voxel grids. In [4] and [5], the principle of random MAP perturbations is introduced. These perturbations allow sampling from the Gibbs distribution, which is inherently defined in CRF models or equivalent energy minimization approaches.

In this work, we extend a typical framework for brain tumor segmentation [6], which is composed of a probabilistic

*These authors contributed equally.

†Institute for Advanced Study, funded by German Excellence Initiative.

local classification scheme and a CRF model, with random MAP perturbations in order to sample brain tumor segmentations. With this approach, we are able to go beyond the typical hard segmentation, ie. the MAP solution, and quantify the uncertainty of the segmentation without additional model assumptions. To our knowledge, we are the first to employ this theoretical framework to obtain uncertainties in medical image segmentation.

2. METHODS

In the following, we briefly review the background on CRFs and the recent work on perturbation models.

The presented work is an extension to the framework of [6]. Starting from a CRF that simultaneously segments brain tumors visible in several standard MR modalities such as T1, T1c and FLAIR, we make use of the perturbation model to quantify the uncertainties in the segmentations.

2.1. CRFs and perturbation models

The CRF is defined over a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $i \in \mathcal{V}$ for each voxel and edges $(i,j) \in \mathcal{E}$ between related vertices. In the particular model that we employ, edges will be introduced for all neighbouring voxels and between the corresponding voxels in all modalities. A variable vector $\mathbf{y}$ describes the assigned labels for all voxels. The energy function $E(\mathbf{y})$ of the CRF is then defined over the graph $\mathcal{G}$ as:

$$E(\mathbf{y}) = \sum_{i \in \mathcal{V}} \phi_i\left(y_i\right) + \sum_{(i,j) \in \mathcal{E}} \phi_{ij}\left(y_i, y_j\right) \ , \tag{1}$$

where $\phi_i\left(y_i\right)$ and $\phi_{ij}\left(y_i, y_j\right)$ are the unary and pairwise potentials, respectively. It is modelled such that favourable states of $\mathbf{y}$ yield low energies. For many practically relevant cases, the energy function can be minimized (approximately) with efficient optimizers such as graph-cut algorithms [7]. For this study, we use the same potential functions as in [6] and the graph-cut algorithm of [8]. However, the perturbation approach is not restricted to this particular choice of potentials.

It is important to notice that the CRF implicitely defines a probability distribution over the segmentations $\mathbf{y}$ in terms of a Gibbs distribution:

$$P\left(\mathbf{y}\right) = \frac{1}{Z} \exp\big(-E(\mathbf{y})\big) \ , \tag{2}$$

where $Z = \sum_{\mathbf{y} \in \mathcal{Y}} \exp\left(-E(\mathbf{y})\right)$ is the partition function that ensures that the distribution sums to 1. This gives us directly a way of quantifying the uncertainty in our segmentation by computing the marginal distributions. Unfortunately, calculating these marginals is intractable in loopy graphs [9] and sampling with MCMC approaches is computationally expensive. To this end, recent work of Papandreou and Yuille [5] has shown that the Gibbs distribution can be approximated by perturbing its energy function with Gumbel noise and solving for its MAP state repeatedly (See Sect. 2.1.3). This allows us to leverage powerful solvers for sampling segmentations from the CRF and thereby approximating the marginal distribution for all voxels.

Next, we elaborate on two particular ways to perturb the energy function that we investigate in this work.

2.1.1. Voxel-specific Gumbel perturbations

Originally, the authors of [5] have proven that samples drawn from the perturbation model coincide in the limit case with the Gibbs distribution if the noise γ is drawn IDD from a Gumbel distribution with zero mean for the full state table of $\mathbf{y}$. While it is not feasible to do this in practice – the full state table has exponentially many entries – several studies [4, 5] have empirically shown that applying low-order perturbations of Gumbel noise yields sufficient results. Our first approach is therefore to perturb each of the unary potentials of the CRF with a sample drawn from the zero-mean Gumbel distribution, as illustrated in Fig. 1. For this case, we can write down the whole perturbation as:

$$\gamma(\mathbf{y}) = \sum_{i \in \mathcal{V}} \sum_{k \in \{T, \neg T\}} \gamma_i^k \, \mathbb{1}(y_i = k) \ , \tag{3}$$

where $\mathbb{1}(.)$ is the indicator function and γ_i^k are samples of the Gumbel distribution with zero mean.

2.1.2. Context-sensitive Gumbel perturbations

Inspired by the Swendsen-Wang cluster-specific updates [2], we explore perturbations on a supervoxel scale in order to detect context-sensitive uncertainties. We parcel the voxel grid into supervoxels (of $\approx$ 1ml) using MonoSLIC [10] and draw perturbations γ from the Gumbel distribution for each supervoxel, resulting in identical perturbations for the unary potentials of all voxels within the same supervoxel. We note that doing so violates the assumptions of the original proof in [5] and the resulting segmentation samples do not follow the original Gibbs distribution of the CRF anymore. Instead, we can interpret it as an additional, context-sensitive correlation prior on the supervoxel grid that is added to the original CRF.

2.1.3. Sampling with perturbation models

Having defined both CRF and the type of perturbation, we can draw samples $\mathbf{y}$ of segmentations with the following procedure:

1. Draw random perturbations $\gamma(\mathbf{y})$ as described in Sect. 2.1.1 or 2.1.2.
2. Draw a new sample $\hat{\mathbf{y}}$ by computing the MAP state of the perturbed CRF:

$$\hat{\mathbf{y}} = \underset{\mathbf{y} \in \mathcal{Y}}{\arg\min}\, E(\mathbf{y}) + \gamma(\mathbf{y}) \ . \tag{4}$$

The two steps are repeated until the desired amount of samples has been created to approximate the marginal distributions (also called *soft segmentation* in our setting). Other than MCMC methods, we do not have to throw away initial samples as there is no burn-in period.

3. EXPERIMENTS AND RESULTS

We conducted experiments on 15 patient-specific datasets, each dataset consisting of 3 time points, and each time point containing 3 standard MR modalities (T1, T1c and FLAIR) as illustrated in Fig. 2. In all T1c and FLAIR volumes, indicating active tumor core and whole tumor similar to the BRATS standards [11], 3-dimensional ground truth annotations have been acquired by a clinical expert using a interactive segmentation tool (SmartBrush, by Brainlab).

Note that this dataset poses a complicated segmentation task. The data contains pre- and post-operative scans depicting resection cavities, internal bleedings and scar tissue.

The datasets are preprocessed by intra-subject registrations, skull extraction and isotropic resampling. We calculated initial brain tumor regions using a generative model as in [12]. Inference is calculated on all modalities at once, as in [6], but time points are processed individually. Using the perturbation models, we sampled 100 segmentations for each time point for each patient.

3.1. Quantitative evaluation

Based on the marginal distributions computed by the voxel-specific perturbation model, we calculated precision-recall curves for each segmentation task. Figure 3 illustrates the confidence bands of the mean precision-recall curve for all T1c and FLAIR segmentation tasks (ground truth was not available for T1), together with the confidence intervals of precision and recall scores for the mean MAP solution of all segmentation tasks. The mean area under curve (AUC) is equal to 71.9% for T1c and 77.9% for FLAIR. For the soft segmentations acquired by the supervoxel-specific perturbation model, the mean AUC is equal to 71.6% for T1c and 70.1% for FLAIR, indicating an overall decrease in segmentation accuracy, especially in FLAIR.

3.2. Qualitative evaluation

Soft segmentation maps resulting from the voxel-specific perturbation model are illustrated in Fig. 2 for one patient-specific dataset. Figure 4 further illustrates the benefit of soft segmentation maps for a few isolated volumes.

Furthermore, we extended the voxel-specific segmentation uncertainties towards volumetric uncertainties. Figure 6 illustrates T1c and FLAIR segmentations (from two different patients) over 3 time points, together with the tumor volumetry uncertainties. The T1c segmentations show a low variation

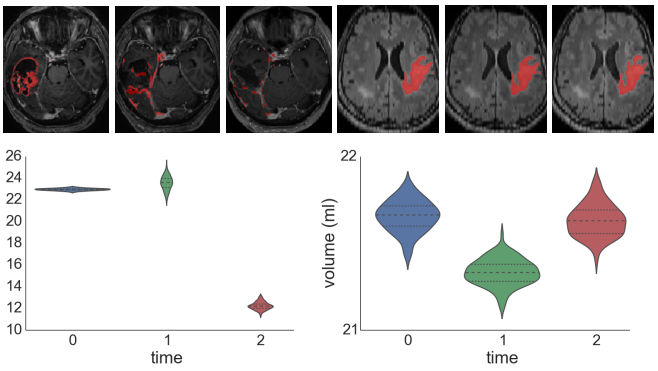

Fig. 6. Uncertainty in brain tumor volumetry. *Left:* T1c tumor segmentations depicting similar volumes for time point 1 and 2, but different levels of uncertainty. *Right:* FLAIR tumor segmentations with high volumetric uncertainties along time.

in volumes across samples for the first time point (where a very clear active tumor rim is visible), while the second time point shows quite a high variation (where the tumor was resected and false positives are present in scar tissue and vessels). The third time point depicts a very clear decrease in active tumor core volume. The FLAIR segmentations show quite high volumetric uncertainties along time, which might be related to the smooth tumor boundaries.

Soft segmentations resulting from supervoxel-specific perturbations are illustrated in Fig. 5. They generally depict larger areas of uncertainty (lower and upper row in Fig. 5) and, in some cases, reflect the underlying structure of the input data better (middle row in Fig. 5).

4. DISCUSSION AND CONCLUSION

We have demonstrated the benefits of probabilistic segmentations in a CRF framework for brain tumor segmentation. Recent work on perturbation models was shown to be well-suited for obtaining samples of segmentations from the CRF model. It introduces minimal overhead and can be applied to virtually all segmentation approaches that rely on graphical models and energy minimization schemes.

5. REFERENCES

[1] A.C. Fan, J.W. Fisher, W.M. Wells, et al., "MCMC curve sampling for image segmentation," in *MICCAI 2007*, vol. 4792 of *LNCS*, pp. 477–485. 2007.

[2] A. Barbu and S.-C. Zhu, "Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities," *IEEE TPAMI*, vol. 27, no. 8, pp. 1239–1253, Aug 2005.

[3] P. Kohli and P.H.S. Torr, "Measuring uncertainty in graph cut solutions," *CVIU*, vol. 112, no. 1, pp. 30–38, Oct 2008.

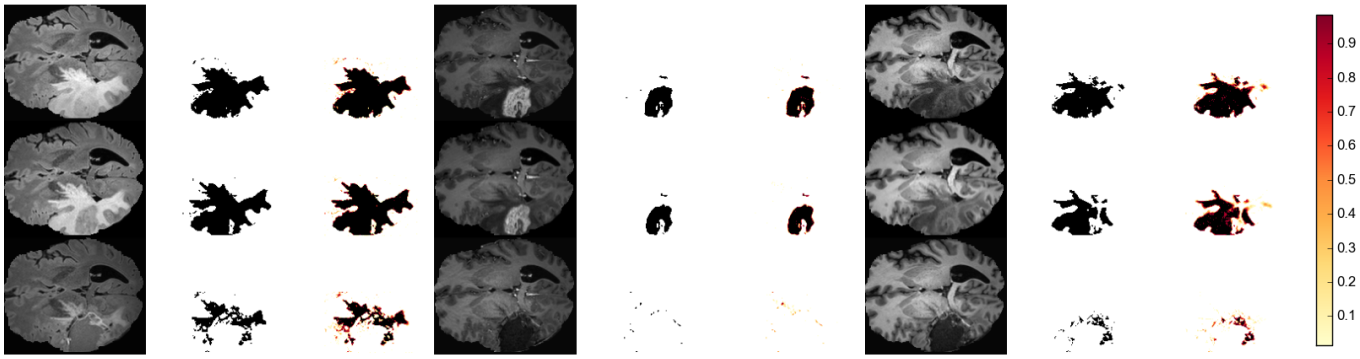

Fig. 2. Patient-specific dataset containing FLAIR, T1c and T1 modalities (*left to right*) for 3 time points (*top to bottom*). For each modality, input data (*left*), hard segmentations, i.e. MAP solutions, (*middle*) and soft segmentations, i.e. marginals, resulting from voxel-specific perturbations (*right*) are shown.

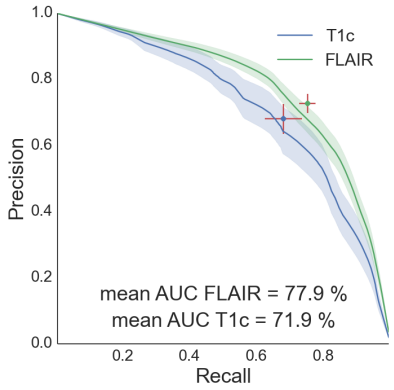

Fig. 3. 68% confidence intervals of the mean precision-recall curve (*blue* and *green*) and the mean MAP solution (*red*).

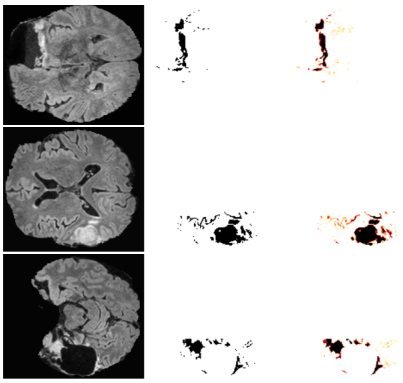

Fig. 4. *Left to right*: input data, hard segmentations (MAP solutions) and soft segmentations.

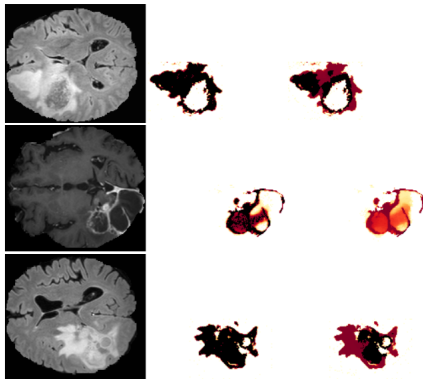

Fig. 5. *Left to right*: input data and soft segmentations resulting from voxel- and supervoxel-specific perturbations.

[4] T. Hazan, S. Maji, and T. Jaakkola, "On sampling from the Gibbs distribution with random maximum a-posteriori perturbations," in *NIPS 26*, pp. 1268–1276. 2013.

[5] G. Papandreou and A.L. Yuille, "Perturb-and-MAP random fields: Using discrete optimization to learn and sample from energy models," in *IEEE ICCV*, pp. 193–200. Nov 2011.

[6] E. Alberts, G. Charpiat, Y. Tarabalka, et al., "A non-parametric model for brain tumor segmentation and volumetry in longitudinal MR sequences," *Proc MICCAI BrainLes (Brain Lesion Workshop)*, p. 8, 2015.

[7] J.H. Kappes, B. Andres, F.A. Hamprecht, et al., "A comparative study of modern inference techniques for structured discrete energy minimization problems," *Int J Comput Vision*, vol. 115, no. 2, pp. 155–184, 2015.

[8] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE TPAMI*, vol. 26, pp. 359–374, 2001.

[9] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, Number 4. MIT press, 2009.

[10] M. Holzer and R. Donner, "Over-segmentation of 3D medical image volumes based on monogenic cues," in *Proceedings of the 19th CVWW*, 2014, pp. 35–42.

[11] B.H. Menze, A. Jakab, S. Bauer, et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE TMI*, p. 33, 2014.

[12] B.H. Menze, K. van Leemput, D. Lashkari, et al., "A generative model for brain tumor segmentation in multi-modal images," in *MICCAI 2010*, vol. 6362 of *LNCS*, pp. 151–159. 2010.