SPATIALLY ADAPTIVE RANDOM FORESTS

Ezequiel Geremia ⋆ *Bjoern H. Menze* ⋆† *Nicholas Ayache* ⋆

⋆ Asclepios Research Project, Inria Sophia-Antipolis, France.
† Computer Vision Laboratory, ETH Zurich, Switzerland.

ABSTRACT

Medical imaging protocols produce large amounts of multi-modal volumetric images. The large size of the datasets contributes to the success of supervised discriminative methods for semantic image segmentation. Classifying relevant structures in medical images is challenging due to *(a)* the large size of data volumes, and *(b)* the severe class overlap in the feature space. Subsampling the training data addresses the first issue at the cost of discarding potentially useful image information. Increasing feature dimensionality addresses the second but requires dense sampling. We propose a general and efficient solution to these problems. "Spatially Adaptive Random Forests" (SARF) is a supervised learning algorithm. SARF aims at automatic semantic labelling of large medical volumes. During training, it learns the optimal image sampling associated to the classification task. During testing, the algorithm quickly handles the background and focuses challenging image regions to refine the classification. SARF demonstrated top performance in the context of multi-class gliomas segmentation in multi-modal MR images.

***Index Terms*—** random forest, multi-scale, hierarchical, structured labelling, sampling, segmentation

1. INTRODUCTION

Medical imaging protocols produce large amounts of multi-modal volumetric images. The large size of the datasets contributes to the success of supervised discriminative methods for semantic label extraction. Automatic classification of semantically relevant structures in medical images is challenging due to *(a)* the large size of data volumes, and *(b)* the severe class overlap in the feature space. Our study focuses on multi-scale segmentation methods which address these issues.

Using multi-scale image representations, relying for example on spectral representations [1], helps speeding up segmentation algorithms and apply them to data sets that were previously considered to be of prohibitive . Interestingly, adaptive multi-resolution hierarchies have also been shown to efficiently encode image information for compression and rendering [2]. In medical applications, recent work focused learning hierarchical anatomical representations explicitly from expert annotations. Alternatively, a generative model can be learnt from expert labelled ground truth and incorporated to a multi-level affinity-based segmentation [3]. The approach in [4] builds on a hierarchical registration and weighting scheme to retrieve organ-specific atlases at different scales. Although, these methods can be adapted to take into account additional image channels, they exclude the use of high dimensional features.

Hierarchical representations can be integrated into discriminative supervised learning algorithms. For instance, boosting weak classifiers that are hierarchically trained on different scales showed to significantly reduce training time [5]. A similar approach was proposed for the segmentation of multiple sclerosis lesions in multi-channel MRIs using random forests [6]. In the latter, segmentations from multiple scales are merged to increase the robustness of the algorithm. In both methods, the image is relabelled at coarse scales to discard regions containing heterogeneous label distributions. As a result, the classifiers miss critical coarse-scale cues which penalizes the performance of the final segmentation. Other supervised approaches build on context-rich random forests for segmentation of brain lesions in MRIs [7, 8]. Still, their excellent performance requires a careful tuning of class weights during training, and is highly sensitive to the spatial sub-sampling of the training data for the different classes.

We present the novel "Spatially Adaptive Random Forest" (SARF) to address these shortcomings. SARF is a supervised learning algorithm which aims at automatic semantic label extraction in multi-modal medical images. It builds on discriminative random forests, an efficient multi-scale 3D image representation, and structured labelling. SARF learns the optimal image sampling associated to the segmentation task from the training data. The ground truth, which is provided at the voxel level, is extrapolated to coarse levels by using label histograms. During both training and testing, the algorithm quickly handles background regions of the image, and focuses on more challenging ones to refine the segmentation. This is made possible by adding a scale transition condition to the random forest algorithm.

We demonstrate SARF in the context of multi-class glioma segmentation in multi-modal MR images. SARF ranked in the top three when applied to the publicly available MICCAI 2012 BRATS Challenge dataset.

2. DATA REPRESENTATION

We derive a hierarchical data representation to efficiently browse the image, and its associated ground truth, at different scales. Fig. 1 illustrates the visual results obtained for a representative case in the BRATS glioma dataset.

2.1. Multi-scale image tree

A multi-scale hierarchical tree is presented for encoding multi-channel volumetric images. It builds on the volumetric counterpart of SLIC superpixels [9] to iteratively generate a compressed representation of the image at different scales. Similarly to spatial trees presented in [10], each layer of the final tree embeds a different scale of the image. The generation of the multi-scale data representation is fast and does not require any parameter tuning. It inherits the strengths of the SLIC algorithm which showed outstanding performance compared to the state-of-the-art [9].

Multi-channel images are defined as the vectorial map $\mathbf{I} : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}^{|\Gamma|}$ where Γ is the set of image channels. Thus, every voxel is associated a spatial position $\mathbf{x} \in \Omega$ and a multi-channel intensity vector $\mathbf{I}(\mathbf{x})$. A spatial partition is a set of disjoint supervoxels $\mathcal{P}_\mathbf{I}^k = \{v_n^k\} \subset \mathcal{P}(\Omega)$ partitioning the image domain Ω at scale k. As presented in [9], the generation of $\mathcal{P}_\mathbf{I}^k$ is exclusively based on the image information $\mathbf{I}$. At the finest scale, $\mathcal{P}_\mathbf{I}^1$ is the set of singletons formed by the individual image voxels. We adapt the SLIC algorithm to recursively cluster supervoxels from $\mathcal{P}_\mathbf{I}^k$ into a coarser partition noted $\mathcal{P}_\mathbf{I}^{k+1}$. In practice, a scalar value s^k controls the maximum size of supervoxels at scale k.

This procedure is repeated for an increasing sequence of supervoxel sizes $\{s^k\}_{k \leqslant K}$ until the whole image is contained in a single supervoxel. In the rest of the article, we set $K = 7$ and $\{s^k\}_{k \leqslant 7}$ to a fixed set of scalars. The resulting sequence $\{\mathcal{P}_\mathbf{I}^k\}$ is encoded in the layers of a multi-resolution tree noted $\mathcal{M}_\mathbf{I} = \{\mathcal{M}_\mathbf{I}^k\}$. At layer $k > 1$, $\mathcal{M}_\mathbf{I}^k$ maps every supervoxel $v \in \mathcal{P}_\mathbf{I}^k$ to a set of disjoint finer supervoxels $\mathcal{M}_\mathbf{I}^k(v) = \{w_i\} \subset \mathcal{P}_\mathbf{I}^{k-1}$ satisfying $v = \bigcup_i w_i$.

2.2. Visual features

Supervoxels provide a convenient primitive from which to compute local image features. A visual feature at scale k is a map defined by $\theta^k : \mathcal{P}_\mathbf{I}^k \rightarrow \mathbb{R}$. Arbitrary large amounts of task-specific features can be derived in a straightforward way from the multi-scale representation of the image [7, 8, 6].

Here, we provide three examples of possible features. A local feature $\theta_{med}^{k,\gamma}$ which maps the supervoxel v to the median intensity in channel γ of the voxels it contains, noted $\theta_{med}^{k,\gamma}(v) = \mu_{\frac{1}{2}}\{\mathbf{I}_\gamma(\mathbf{x}) | \mathbf{x} \in v\}$. A prior feature $\theta_{prior}^{k,\delta} = \theta_{med}^{k,\delta}$ where the channel δ maps the spatial distribution of healthy tissues including white matter (WM), grey matter (GM) and cerebro-spinal fluid (CSF). The prior feature is obtained by affinely registering the MNI atlases on each patient. A long-range feature defined by $\theta_{sym}^{k,\gamma}(v) = \max\{S \circ \mathbf{I}_\gamma(\mathbf{x}) | \mathbf{x} \in v\}$, where S is a reflection of $\mathbb{R}^3$. The symmetry feature was designed detecting abnormal regions in brain MRIs which are often asymmetrical with respect to the mid-sagittal plane.

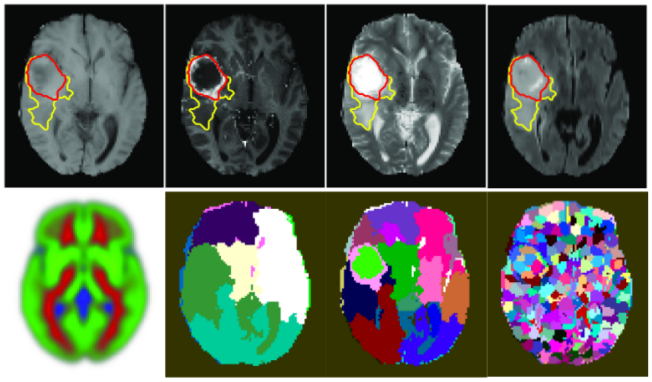


Fig. 1. **Data representation.** Top: the multi-channel MRI noted $\mathbf{I}$, including (T1, T1+gadolinium, T2, FLAIR) overlayed with the expert annotations for the edema (yellow) and the tumor core (red). Bottom left: the affinely registered MNI atlas (WM, GM, CSF), and the image partitions at three different scales ($\mathcal{P}_\mathbf{I}^2$,$\mathcal{P}_\mathbf{I}^4$,$\mathcal{P}_\mathbf{I}^6$).

2.3. Ground truth

At the voxel level, the ground truth associated to the image $\mathbf{I}$ is defined by $G_\mathbf{I} : \Omega \rightarrow \mathcal{C}$, each voxel $\mathbf{x}$ being associated a class label $G_\mathbf{I}(\mathbf{x}) \in \mathcal{C}$. Here, we consider the tissue classes $\mathcal{C} = \{back, edema, core\}$ standing for background healthy brain, edema and tumor core, respectively. The generalization of $G_\mathbf{I}$ to coarser scales reads $\mathcal{G}_\mathbf{I} = \{\mathcal{G}_\mathbf{I}^k\}$ where $\mathcal{G}_\mathbf{I}^k : \mathcal{P}_\mathbf{I}^k \rightarrow \mathcal{C}$.

In previous work [5, 6], each supervoxel v was affected the class label $\mathcal{G}_\mathbf{I}^k(v) = c \in \mathcal{C}$ satisfying $|\{\mathbf{x} \in v \mid G_\mathbf{I}(\mathbf{x}) = c\}| > \tau_{hom}$. When such label did not exist, the supervoxel was removed from the training set. The threshold $\tau_{hom} = 70$ or 80% aims at selecting homogeneous supervoxels, while discarding those showing severe label mixture. In multi-class segmentation, this means discarding challenging, but often critical, image regions, and thus indirectly penalizing the prediction performance.

To address this flaw, we introduce an unambiguous labelling function $\mathcal{H}_\mathbf{I}^k$ with values in $\mathbb{N}^{|\mathcal{C}|}$. The histogram $\mathcal{H}_\mathbf{I}^k(v) = (h[c])_{c \in \mathcal{C}} = (|\{\mathbf{x} \in v \mid G_\mathbf{I}(\mathbf{x}) = c\}|)_{c \in \mathcal{C}}$ counts the class label occurrences in the supervoxel v. Consequently at scale k, the ground truth is defined as $\mathcal{G}_\mathbf{I}^k(v) = \arg\max_{c \in \mathcal{C}} h[c]$. Unlike [5, 6], our labelling method $\mathcal{H}_\mathbf{I} = \{\mathcal{H}_\mathbf{I}^k\}$ keeps track of the class mixture in every supervoxel. In Section 3.3, we explain how this is integrated to the random forest framework to help refining the segmentation in challenging image regions.

3. SPATIALLY ADAPTIVE RANDOM FOREST

The random forest framework [11] is extended to benefit from the presented multi-scale image representation. In the following, we provide a sound formulation of SARF applicable to the general problem of multi-class image segmentation.

3.1. Training data

The SARF is an ensemble of trees, each processing the multi-scale data hierarchy from coarse to fine. During training, the data entering the root node of each tree consists of all supervoxels $v_j \in \bigcup_n \mathcal{P}^K_{\mathbf{I}_n}$, where n indexes the case, considered at the coarsest scale K. In the following, $\mathcal{M}$, $\mathcal{G}$ and $\mathcal{H}$, denote the extensions to the dataset $\{\mathbf{I}_n\}$ of $\mathcal{M}_{\mathbf{I}}$, $\mathcal{G}_{\mathbf{I}}$, and $\mathcal{H}_{\mathbf{I}}$, respectively. Every supervoxel v_j is associated the class label $c_j = \mathcal{G}^k(v_j)$, and the label histogram $h_j = \mathcal{H}^k(v_j)$. The resulting training data reads $\mathcal{T} = \{(v_j, c_j, h_j)\}$ with $j \in J$.

3.2. Decision node representation

Each internal node p applies a binary test $t^{k,\theta,\tau}_{J_p}$ to the data it receives $\mathcal{T}^k_p = \{(v_j, h_j)\}_{j \in J_p}$. The binary test is defined by $t^{k,\theta,\tau}_p(v_j) = \theta^k(v_j) > \tau$. It is parametrized by the scale k, the type of visual feature θ and the threshold τ. Based on the outcome of this test, $\mathcal{T}^k_p$ is splitted into $\mathcal{T}^k_{L(p)}$ and $\mathcal{T}^k_{R(p)}$, which are propagated to the left and right child node receptively.

During training, every node p stores the scale k, the optimal parameters θ and τ used to split the data. Additionally, it saves the label distribution $d_{\mathcal{T}^k_p} = (\,|\{j \in J_p \mid c_j = c\}|\,)_{c \in \mathcal{C}}$, and the class mixture $h_{\mathcal{T}^k_p} = \sum_{j \in J_p} h_j$ considered on $\mathcal{T}^k_p$.

3.3. Training

For each node p, the parameter $\lambda = (\theta, \tau)$ is optimized using the input training data $\mathcal{T}^k_p$. The optimality criterium is the information gain defined as $IG(\lambda, \mathcal{T}^k_p) = H(d_{\mathcal{T}^k_p}) - \sum_{B \in \{L,R\}} w_B \cdot H(d_{\mathcal{T}^k_{B(p)}})$, where $w_B = |\mathcal{T}^k_{B(p)}|/|\mathcal{T}^k_p|$. The entropy H satisfies $H(d_{\mathcal{T}^k_p}) = -\sum_{c \in \mathcal{C}} P(c) \cdot log\, P(c)$ with $P(c) = d_{\mathcal{T}^k_p}[c] / \sum_{c' \in \mathcal{C}} d_{\mathcal{T}^k_p}[c']$. The optimal parameters satisfy $\lambda^*_p = \arg\max_\lambda IG(\lambda, \mathcal{T}^k_p)$.

This procedure is repeated recursively for the derived nodes. The tree is grown down to scale $k = 1$ until every leaf node p is pure, i.e. $H(\mathcal{T}^1_p) = 0$. Unlike previous work [6, 7, 8], we introduce spatial refinement in the random forest framework to capture fine structures. Indeed, when the supervoxels are too large to properly describe annotated image regions, the scale k is decremented. Formally, this occurs when $H(d_{\mathcal{T}^k_p}) = 0$ and $H(h_{\mathcal{T}^1_p}) \neq 0$. In this case, $\mathcal{T}^k_p = \{(v_j, h_j)\}_{j \in J_p}$ is replaced by its decomposition into finer supervoxels $\mathcal{T}^{k-1}_p = \{(w_i, c'_i, h'_i)\}_{i \in I_p}$ where $w_i \in \bigcup_j \mathcal{M}^k(v_j)$, $c'_i = \mathcal{G}^{k-1}(w_j)$, and $h'_i = \mathcal{H}^{k-1}(w_j)$. The node is then optimized at the scale $k - 1$.

We solve the optimization problem by exhaustive search over a random set of thresholds and the whole set of feature $\{\theta_{med}, \theta_{prior}, \theta_{sym}\}$. To further decorrelate the weak classifiers, we train each tree with a different partition of the same image. This is done by randomizing the initial position of the seeds in the SLIC algorithm [9].

3.4. Prediction

When applied to an unseen test volume $\mathbf{I}_{test}$, a different image partition $\mathcal{P}^{K,t}_{\mathbf{I}_{test}}$ is computed for every tree $t \in [1..T]$. In every tree, each node p applies the binary test t^{k,θ^*,τ^*} to the input data, after having refined it to scale k, if necessary. As a result, for every voxel $v \in \mathcal{P}^{1,t}_{\mathbf{I}_{test}}$ there is a tree-specific sequence of supervoxels $\{v^k_t\}_{k \in [k_{min}, K]}$ of decreasing size such that $v^k_t \in \mathcal{P}^{k,t}_{\mathbf{I}_{test}}$. The finest supervoxel $w_t = v^{k_{min}}_t$ reaches the leaf node p_t, and is affected the associated posterior class distribution $d_t(w_t) = d_{\mathcal{T}^k_{p_t}} / \sum_{c \in \mathcal{C}} d_{\mathcal{T}^k_{p_t}}[c] \in \mathcal{L}$. For every voxel $v \in \mathcal{P}^1_{\mathbf{I}_{test}}$, posteriors from all trees are averaged to from the forest posterior such that $d_f(v) = \sum_t d_t(w_t)/T$. Finally, the predicted class label affected to v is $\widehat{G}_{\mathbf{I}_{test}}(v) = \arg\max_{c \in \mathcal{C}} d_f(v)[c]$.

4. EXPERIMENTS AND RESULTS

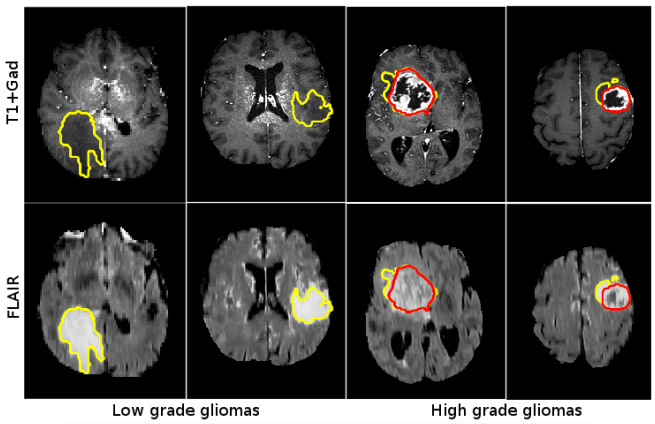
We demonstrate SARF in the specific context of multi-class glioma segmentation in multi-modal MRIs. For training purposes, we rely on the publicly available MICCAI BRATS Challenge 2012 (MBC) dataset which contains 80 cases. [1] Our method was then applied to a test dataset of 30 cases which ground truth was kept secret. After independent evaluation of the results by the challenge website, SARF ranked third among eight state-of-the-art methods. Additionally, SARF performed 90% faster than classical random forests applied without sub-sampling for comparable results. Obtained results are illustrated in Fig. 2.

5. DISCUSSION AND CONCLUSION

The presented SARF framework shows promising results still limited by the small size of the employed feature set. Indeed, it would greatly benefit from richer visual features designed for glioma segmentation, and subsequent regularization as used in top-ranked methods [8, 12].

Fig. 3 shows that the SARF focuses on challenging image regions by processing them at finer scales. Interestingly, SARF automatically finds the average optimal scale used to

[1] Brain tumor image data used in this work were obtained from the MICCAI 2012 Challenge on Multimodal Brain Tumor Segmentation (http://www.imm.dtu.dk/projects/BRATS2012) organized by B. Menze, A. Jakab, S. Bauer, M. Reyes, M. Prastawa, and K. Van Leemput. The challenge database contains fully anonymized images from the following institutions: ETH Zurich, University of Bern, University of Debrecen, and University of Utah.



Method	Dice score (edema)	Dice score (core)
[8]	0.67	0.65
[12]	0.61	0.63
SARF	0.62	0.55

Fig. 2. **MBC challenge results.** FLAIR and T1+Gad images overlayed with the predicted edema (yellow) and tumor core (red). Below the Dice scores of the top-ranked methods for the edema and tumor core classes.

segment gliomas, here $k = 3$. This supports the fact that parsing the image at the voxel level is often unnecessary. These findings arise from the novelty of the multi-scale classification approach implemented in SARF, and result in significantly reducing training and testing times.

Aknowledgment

Part of this work was funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080.

6. REFERENCES

[1] T. Cour, F. Bénézit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," in *CVPR (2)*, 2005.

[2] S. Lefebvre and H. Hoppe, "Compressed random-access trees for spatially coherent data," in *Rendering Techniques*, 2007.

[3] J. J. Corso, E. Sharon, S. D., S. El-Saden, Usha Sinha, and Alan L. Yuille, "Efficient multilevel brain tumor segmentation with integrated bayesian model classification," *IEEE Trans. Med. Imag.*, vol. 27, 2008.

[4] R. Wolz, C. Chu, K. Misawa, K. Mori, and D. Rueckert, "Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases," in *MICCAI (1)*, 2012.

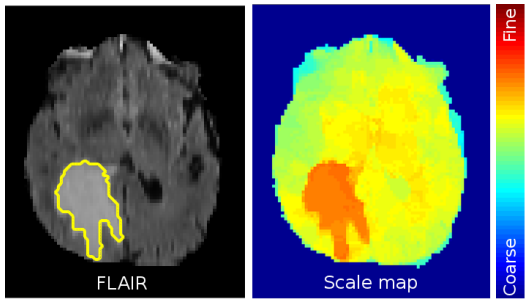


Fig. 3. **Scale map.** FLAIR image with overlayed ground truth (yellow = edema), and the corresponding scale map displaying for each voxel the scale used to classify it (dark red for $k = 1$ and dark blue for $k = K = 7$).

[5] J. A. dos Santos, P. H. Gosselin, S. Philipp-Foliguet, R. da Silva Torres, and A. X. Falcão, "Multiscale classification of remote sensing images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 50, no. 10, 2012.

[6] A. Akselrod-Ballin, M. Galun, M. John Gomori, M. Filippi, P. Valsasina, R. Basri, and A. Brandt, "Automatic segmentation and classification of multiple sclerosis in multichannel MRI," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 10, 2009.

[7] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache, "Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images," *NeuroImage*, vol. 57, no. 2, 2011.

[8] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. M. Thomas, T. Das, R. Jena, and S. J. Price, "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," in *MICCAI (3)*, 2012.

[9] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, 2012.

[10] J. M. Siskind, J. Sherman Jr., I. Pollak, M. P. Harper, and C. A. Bouman, "Spatial random tree grammars for modeling hierarchal structure in images with regions of arbitrary shape," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, 2007.

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, 2001.

[12] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *MICCAI (3)*, 2011.