

# Visual Saliency Based Active Learning for Prostate MRI Segmentation

Dwarikanath Mahapatra<sup>(✉)</sup> and Joachim M. Buhmann

Department of Computer Science, ETH Zurich, Zürich, Switzerland  
mahapatd@vision.ee.ethz.ch

**Abstract.** We propose an active learning (AL) approach for prostate segmentation from magnetic resonance (MR) images. Our label query strategy is inspired from the principles of visual saliency that has similar considerations for choosing the most salient region. These similarities are encoded in a graph using classification maps and low level features. Random walks identify the most informative node which is equivalent to the label query sample in AL. Experimental results on the MICCAI 2012 Prostate segmentation challenge show the superior performance of our approach to conventional methods using fully supervised learning.

## 1 Introduction

According to the American Cancer society, prostate cancer is the second leading cause of cancer death in American men, and early diagnosis can potentially increase the survival rate amongst patients [1]. Accurate quantification of prostate volume (PV), and location relative to adjacent organs is also an essential part of image guided radiation therapy (IGRT). Magnetic resonance imaging's (MRI) popularity in treatment planning has increased due to high spatial resolution, soft-tissue contrast and absence of ionising radiations.

Manual segmentation of the prostate in MRI is time consuming, and prone to inter- and intra-expert variability. This necessitates the design of (semi-) automated segmentation algorithms that can overcome challenges like: 1) variability of prostate size and shape between subjects; 2) variable image appearance and intensity ranges from different MR scanning protocols; and 3) lack of clear prostate boundaries due to similar intensity profiles of surrounding tissues. Hence, machine learning (ML) methods have focused on learning discriminative image features from manual annotations. However, manual annotations are very expensive, time consuming and requires personnel with high expertise.

The growing importance of prostate MRI segmentation led to the prostate segmentation challenge in MICCAI 2012. Different approaches in the challenge include marginal space learning [2], multi-atlas segmentation [7] and ML [8]. Success of ML methods depends on the discriminative power of hand crafted features. To overcome this shortcoming Liao et al. in [11] propose a deep learning framework using independent subspace analysis (ISA) to automatically learn the most discriminative features. A detailed overview of the different methods and their performance is found in [12].

In this paper we propose an active learning based segmentation method that requires significantly fewer labeled samples, yet achieves higher segmentation accuracy than conventional ML methods. The important contribution of this paper is a visual saliency based approach to select the most informative samples for active learning (AL). We show that many of the principles of salient image region detection are applicable to query selection in active learning. Hence selecting the most informative region in MRI becomes a problem of salient region detection by defining an appropriate measure of a region's importance.

## 2 Image Features

We calculate the mean, variance, skewness and kurtosis of intensity, texture and mean 3D curvature values from a  $31 \times 31$  patch around every voxel. The texture maps are calculated for each slice of the supervoxel using 2D Gabor filters oriented at  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  at the original scale. Thus each voxel gives a 24 dimensional feature vector.

## 3 Semi Supervised Learning With Random Forests

Random forests (RF) [3] have become increasingly popular in classification tasks because of their computational efficiency for large training data and ability to handle multiclass classification. Semi supervised learning techniques train classifiers with a few labeled samples [4] and many unlabeled samples. This is a typical scenario in many medical applications where it is difficult to find qualified experts to label the large number of medical images. A 'single shot' RF method for SSL without the need for iterative retraining was introduced in [6]. We use this method for SSL as it is shown to outperform other approaches.

For labeled samples the information gain over data splits at each node is maximised and encourages separation of the labeled data [3, 6]. However for SSL the objective function encourages separation of the labeled training data and simultaneously separates different high density regions. It is achieved via the following mixed information gain:

$$I_j = I_j^U + \beta I_j^S \quad (1)$$

where  $I_j^S = H(S_j) - \sum_{i \in \{L, R\}} \frac{|S_j^i|}{|S_j|} H(S_j^i)$  is the information gain from the labeled data;  $H$  is the entropy of training points,  $S_j^L$  and  $S_j^R$  the subsets going to the left and right children of node  $j$ , and  $\beta$  is a user defined weight.  $I_j^U$  depends on both labeled and unlabeled data, and is defined using differential entropies over continuous parameters as

$$I_j^U = \log |\Lambda(S_j)| - \sum_{i \in \{L, R\}} \frac{|S_j^i|}{|S_j|} \log |\Lambda(S_j^i)| \quad (2)$$

$\Lambda$  is the covariance matrix of the assumed multivariate distributions at each node. Further details are given in [6].

## 4 SSL-AL Based Segmentation From MR Images

**Initial Preprocessing.** The given images are first bias-corrected using the method in [5] and the intensities are then to  $[0, 1]$ . Learning starts with randomly chosen labeled samples (voxels) of the first training dataset (set  $L$ ). The features of  $L$  and unlabeled voxels (set  $U$  from the remaining training datasets) are used as inputs to an RF based SSL classifier (denoted as  $RF - SSL$ ) which predicts the class labels and probabilities of the unlabeled patches. The most informative sample is added to set  $L$  and the classifier is updated using online learning [15]. In online learning there is no need for retraining of classifiers. The RF classifier is updated based on the newly labeled samples only. The query strategy of AL is discussed in Section 4.1.

### 4.1 AL Query Strategy

Conventional query strategies like density weighting [16] use classification uncertainty and weigh samples according to their similarity with other neighbors. However conventional approaches do not exploit the contextual information over neighborhoods. Our query strategy selects a sample: 1) with high classification uncertainty to obtain novel information from each labeling instance; 2) situated in a dense region such that it is representative of many other samples; and 3) minimal overlap of influence with previously labeled samples to minimize redundancy in labeling effort. Salient image regions have the following characteristics: 1) feature values are significantly different from surroundings (high local contrast); and 2) contrast magnitude is higher than other regions. High contrast regions have maximum information and hence higher entropy [10]. High classification uncertainty also corresponds to high entropy, indicating a correspondence between information content of salient regions and classification uncertainty. Salient regions are located on regular objects (or dense regions of the sample space) and different salient regions are far away from each other, i.e. their influence areas have minimum overlap. Thus we see that the properties of salient regions have similarities with the desired characteristics of query samples. Hence saliency models can be adapted for active learning tasks using appropriate similarity metrics.

Image patches are represented as nodes  $V$  of a graph  $G$ , and connected by set of edges  $E$ . Based on the similarity between any two nodes  $i$  and  $j$  a weight  $w_{ij}$  is given by

$$w_{ij} = \exp \left( \frac{-\|F_i - F_j\|^2}{\sigma^2} \right), \quad (3)$$

where  $F_i$  is the feature vector of node  $i$  or the informativeness  $Inf(i)$ ;  $\sigma = 1$ , and  $\|\cdot\|$  denotes  $L_2$  norm. Informativeness of node  $i$  (or voxel  $x$ ) is given by

$$Inf(i) = \{\phi(i), \gamma, \alpha\}. \quad (4)$$

$\phi$  is the classification uncertainty of  $i$  given by the entropy as

$$\phi(i) = - \sum_{\hat{y}} P((\hat{y}|i) \log P((\hat{y}|i), \quad (5)$$

where  $\hat{y}$  indicates all possible labels (in this case two) for  $i$ , and  $P((\hat{y}|i)$  is calculated by RF-SSL. High entropy indicates greater uncertainty.  $\alpha$  incorporates contextual information, and  $a$  is the collection of intensity, texture and curvature differences defined as

$$\gamma = [Int_{ij} \quad Tex_{ij} \quad Curv_{ij}]. \quad (6)$$

where  $Int_{ij} = \sum_{j \in N} e^{-|Int_i - Int_j|/\sigma^2}$  is the sum of exponential of intensity differences between node  $i$  and all unlabeled nodes  $j$  in  $N$  (a  $48 \times 48$  neighborhood of  $x$ ), and  $\sigma = 1$ . For similar nodes,  $Int_{ij}$  takes higher values.  $Tex_{ij}$  and  $Curv_{ij}$  are the corresponding texture (from the oriented map at  $90^\circ$ ) and curvature differences. Note that we do not average the feature differences over the neighborhood. In a high density region  $\gamma$  is calculated by summing over more voxels than in a sparsely populated region. Since  $\gamma$  is not divided by a normalization constant its value is higher in a high density region.

**Context Information for Informativeness:** An unlabeled sample close to a labeled sample is assigned lower importance because it has a higher probability of having the same label than a sample far away. If the radiologist were to annotate samples close to an already labeled sample it *does not* generally lead to significant information gain. Thus  $\alpha$  incorporates context information and is equal to  $i$ 's distance from the nearest labeled sample

$$\alpha = \min (\|i - i^L\|). \quad (7)$$

where  $i^L$  denotes all the labeled samples (or nodes), and  $\|\cdot\|$  denotes the Euclidean distance based on voxel co-ordinates.

## 4.2 Random Walks and Most Salient Node

The weights  $w_{ij}$  (Eqn.3) are used to define the affinity matrix  $A$  as

$$A_{ij} = \begin{cases} w_{ij}, i \neq j \\ 0, i = j. \end{cases} \quad (8)$$

The most salient node is identified by the random walks algorithm on the graph. Let us denote as  $E_i(T_j)$  the expected number of steps to reach state  $j$  if a Markov chain is started in state  $i$  at time  $t = 0$ . It is also known as the hitting time, and can be derived from the fundamental matrix ( $\mathbf{Z}$ ) of an ergodic Markov chain and its equilibrium probability distribution  $\pi$ . The global saliency of node  $i$  is given by the sum of hitting times from all other nodes to node  $i$  on a complete graph,

$$H_i = \sum_j E_j(T_i), \quad (9)$$

and the most salient node is given by the maximum  $H_i$  as  $N_s = \arg \max_i H_i$ . For details the reader is referred to [9]. Labels are queried for  $N_s$ .

**Stopping Criteria:** Irrespective of the number of labeled samples, there will always be one unlabeled sample with maximum informativeness. In order to ensure that the label query does not continue indefinitely, we determine the probability values of the two classes for the most informative sample. If the probability value for any one class is less than 0.45 (or greater than 0.55 for the other class) we do not query the label for that sample. If we encounter such samples for 5 consecutive iterations then we stop the label query because this indicates that the classifier has obtained sufficient samples to have high confidence on its classification output.

### 4.3 Graph Cut Segmentation

A spatially smooth solution is obtained by formulating the segmentation as a labeling problem within a second order Markov random field (MRF) cost function. The labels are obtained for each voxel and not the individual patches by optimizing the cost function using graph cuts. A second order MRF energy function is given by

$$E(L) = \sum_{s \in P} D(L_s) + \lambda \sum_{(s,t) \in N} V(L_s, L_t), \quad (10)$$

where  $P$  denotes the set of pixels and  $N$  is the set of neighboring pixels for pixel  $s$ .  $\lambda$  is a weight that determines the relative contribution of penalty cost ( $D$ ) and smoothness cost ( $V$ ).  $D(L_s)$  is given by

$$D(L_s) = -\log (Pr(L_s) + \epsilon), \quad (11)$$

where  $Pr$  is the likelihood (from probability maps) previously obtained using RF classifiers and  $\epsilon = 0.00001$  is a very small value to ensure that the cost is a real number. The penalty cost encourages high label probability.  $V$  ensures a spatially smooth solution by penalizing discontinuities and is defined as

$$V(L_s, L_t) = e^{-\frac{(I_s - I_t)^2}{2\sigma^2}} \cdot \frac{1}{\|s - t\|}, \quad (12)$$

$I$  denotes the intensity. Smoothness cost is determined over a 8 neighborhood system.

## 5 Experiments and Results

We apply our method on the MICCAI 2012 PROMISE prostate segmentation challenge (<http://promise12.grand-challenge.org/>). The training set consisting of 50 patients is used to train our classifier, which is then applied on the 30 test datasets of transversal T2-weighted MR images. We submit our results to the online evaluation system and get feedback on our performance as well as a ranking. The datasets are acquired under different clinical settings.

They are multi-center and multi-vendor, and have different acquisition protocols (e.g. differences in slice thickness, with/without endorectal coil). The volume dimensions and voxel resolutions are different for different images. Each slice of the different volumes is of size  $512 \times 512$  (voxel resolution of  $0.4 \times 0.4 \times 3.3$ ) or  $320 \times 320$  (resolution  $0.6 \times 0.6 \times 3.6$ ). Our whole pipeline was implemented in MATLAB on a 2.66 GHz quad core CPU running Windows 7. The quality of our segmentation results was evaluated by the organizers using: 1) Dice Metric (DM) and 2) 95% Hausdorff Distance (HD).  $\lambda$  (Eqn. 10) was set to 0.02. Our RF-SSL classifier had 50 trees with tree depth of 20. Due to space constraints we provide results only on the test set which is used to rank each algorithm.

### 5.1 Results on MICCAI PROMISE12 Online Challenge Dataset

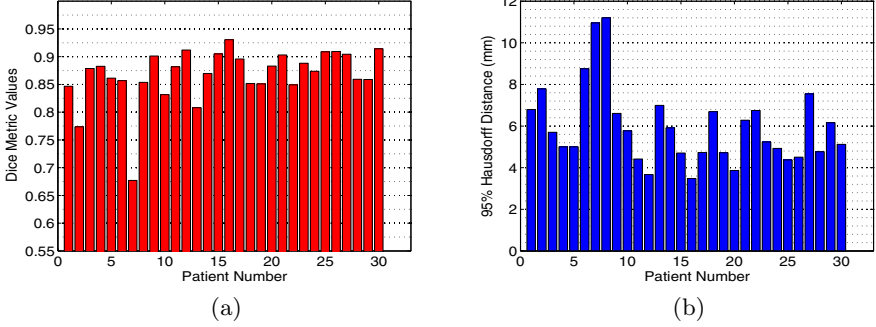
We train our *SSL – AL* classifier on all 50 volumes of the training data and submit our results for 30 challenge datasets for which the manual segmentations are not available to the participants. The ranking methodology is explained in [12]. Our proposed *SSL – AL* method is ranked **third** among all the methods while a competing fully supervised learning (FSL) based method [13] is ranked 14.

*SSL – AL*’s quantitative values are  $DM = 86.7 \pm 4.9$ ,  $95\%HD = 5.9 \pm 1.9$  mm, and boundary distance =  $2.25 \pm 0.77$  mm. The corresponding values for FSL are  $DM = 80.6 \pm 6.5$ ,  $95\%HD = 7.6 \pm 1.7$  mm, and boundary distance =  $3.38 \pm 0.84$  mm. A plot of DM and HD values for our method on individual datasets is shown in Fig. 1.

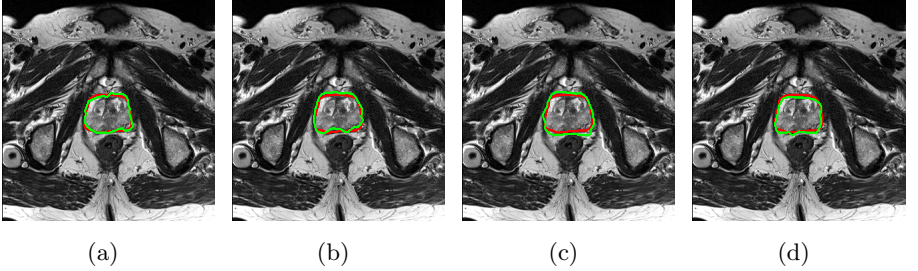
[17]’s method is ranked first followed by [2] with the following scores: 1) [17] -  $DM = 88.0 \pm 4.0$ ,  $95\%HD = 5.94 \pm 2.14$  mm, and boundary distance =  $2.1 \pm 0.68$  mm; and 2) [2] -  $DM = 87.0 \pm 4.0$ ,  $95\%HD = 5.58 \pm 1.49$  mm, and boundary distance =  $2.13 \pm 0.48$  mm. For *DM*, *SSL – AL* ranked fourth with [14] having  $DM = 87.0 \pm 4.0$ , ranked third. For all other metrics our method was ranked third. Importantly, our method’s *DM* and *HD* values are very close to the two methods ranked higher than us. The significant improvement in performance of *SSL – AL* over *FSL* clearly indicates the advantage of using SSL and AL in training a classifier. The difference in their performance is also significant ( $p < 0.0001$ ). Fig. 2 shows segmentation results on Patient 15 from the training data using different methods. Since we do not have access to the manual segmentations of the online challenge dataset, we are unable to show the comparative performance of the difference methods.

### 5.2 Savings in Labeling Effort and Time

By querying labels of most informative patches we reduce the redundancy of labels such that new labels provide truly novel information to the classifier. In 5-fold cross validation *FSL* uses approximately 80% of manual labels for training while *SSL – AL* requires 42% of manual annotations and still achieves higher segmentation accuracy. Although *FSL* has access to more training samples it performs poorly. More training samples does not necessarily translate to better performance since they could introduce noise, particularly if the annotations are



**Fig. 1.** Individual values for  $SSL - AL$  on 30 patients of the MICCAI 2012 online challenge dataset: (a) DM; (b) 95%  $HD$



**Fig. 2.** Segmentation results for Patient 15. The manual annotations are shown in red with the algorithm segmentations in green: (a)  $SSL - AL$  (b)  $FSL$  (c)  $SSL - AL_{n\alpha}$  and (d)  $FSL - AL$ .

not accurate enough. In such a scenario it is beneficial to ask experts to label only the most informative samples. This leads to savings in time, effort and also improves algorithm's performance.

Online learning leads to significant time savings. Updating the classifiers with every new label requires about 0.03–0.1 seconds. On the other hand, lots of time is required to retrain the entire classifier using the entire training set. The total *training time* for the 50 training patients is 38 minutes using  $FSL$  (as reported in [13]) and 20 minutes using  $SSL - AL$ . The lower training time for  $SSL - AL$  is due to fewer training samples, leading to time savings of  $(1 - 20/38) = 47\%$ . However,  $FSL - AL$  takes a longer training time (32 minutes) than  $SSL - AL$  since it does not make use of unlabeled samples. However it is still less than  $FSL$  since many of the manually labeled samples are redundant.

## 6 Discussion and Conclusion

We have developed a novel query strategy for active learning based prostate segmentation. The problem of query sample selection is similar to detecting visually

salient regions on a graph. Our method combines semi supervised classification and active learning to achieve higher segmentation accuracy than fully supervised methods, but with fewer labeled samples. Experimental results on real patient prostate MR volumes from the public MICCAI 2012 Prostate segmentation challenge dataset show our method is ranked third amongst 16 methods. Our performance is quite close to the two methods ranked above us. This clearly demonstrates the improvement in segmentation accuracy obtained using SSL and AL even without knowledge of labels of the test image.

## References

1. Cancer facts and figures (2014). amer. cancer soc. <http://www.cancer.org>
2. Birkbeck, N., et. al.: Region-specific hierarchical segmentation of MR prostate using discriminative learning. In: Proc. MICCAI PROMISE 2012, pp. 4–11 (2012)
3. Breiman, L.: Random forests. *Machine Learning* **45**(1), 5–32 (2001)
4. Chapelle, O., Scholkopf, B., Zien, A.: *Semi-Supervised Learning*. MIT Press, Cambridge (2006)
5. Cohen, M., et al.: Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Human Brain Map* **10**(4), 204–211 (2000)
6. Criminisi, A., Shotton, J.: *Decision Forests for Computer Vision and Medical Image Analysis*. Springer
7. Gao, Q., et. al.: An automatic multi-atlas based prostate segmentation using local appearance-specific atlases and patch-based voxel weighting. In: Proc. MICCAI PROMISE 2012, pp. 12–19 (2012)
8. Ghose, S., et. al.: A random forest based classification approach to prostate segmentation in MRI. In: Proc. MICCAI PROMISE 2012, pp. 20–27 (2012)
9. Harel, J., Koch, C., Perona, P.: Graph based visual saliency. In: NIPS, pp. 545–552 (2006)
10. Kadir, T., Brady, M.: Saliency, scale and image description. *International Journal of Computer Vision* **45**(2), 85–105 (2001)
11. Liao, S., Gao, Y., Oto, A., Shen, D.: Representation learning: a unified deep learning framework for automatic prostate MR segmentation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 254–261. Springer, Heidelberg (2013)
12. Litjens, G., et al.: Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Med. Imag. Anal.* **18**(2), 359–373 (2014)
13. Mahapatra, D., Buhmann, J.: Prostate mri segmentation using learned semantic knowledge and graph cuts. *IEEE Trans. Biomed. Engg.* **61**(3), 756–764 (2014)
14. Malmberg, F., et. al.: Smart paint a new interactive segmentation method applied to MR prostate segmentation. In: Proc. MICCAI PROMISE 2012, pp. 4–11 (2012)
15. Saffari, A., Leistner, C., Santner, J., Godec, M., Bischof, H.: On-line random forests. In: IEEE ICCV Workshops, pp. 1393–1400 (2009)
16. Settles, B., Craven, M.: An analysis of active learning strategies for sequence labeling tasks. In: *Empirical Methods in Natural Language Processing*, pp. 1070–1079 (2008)
17. Vincent, G., et. al.: Fully automatic segmentation of the prostate using active appearance models. In: Proc. MICCAI PROMISE 2012, pp. 75–81 (2012)



Machine Learning in Medical Imaging  
6th International Workshop, MLMI 2015, Held in  
Conjunction with MICCAI 2015, Munich, Germany,  
October 5, 2015, Proceedings  
Zhou, L.; Wang, L.; Wang, Q.; Shi, Y. (Eds.)  
2015, XII, 341 p. 128 illus., Softcover  
ISBN: 978-3-319-24887-5