Machine-learning based image segmentation using Manifold Learning and Random Patch Forests

K. Fritscher[1], P. Raudaschl [1], P. Zaffino[2], G.C. Sharp[3], M.F. Spadea[2], R. Schubert[1]

[1] Department for Biomedical Image Analysis, UMIT, Austria
[2] Department of Experimental and Clinical Medicine, Magna Graecia University, Italy
[3] Massachusetts General Hospital, Harvard Medical School, USA

Abstract. Accurate segmentation of organs at risk is an essential precondition for successful treatment planning in radiotherapy. In this paper we introduce a novel approach for the fully automatic segmentation of organs at risk in CT images.

The method is based on the usage of Manifold Learning in order to extract features of multi-scale image patches. In combination with spatial locality information, the resulting embedding coordinates are used as features for a multi-step classification approach based on using multiple Random Patch Forests.

Experiments using 17 CT images of the head-neck region demonstrate the high potential of the presented method. Quantitative evaluation shows that the approach provides accurate segmentation results even with a small number of training subjects.

1 Introduction

Efficient and exact delineation of anatomical structures is one of the key areas in medical imaging. This is especially true for medical fields like radiation oncology, in which the segmentation of anatomical structures (e.g. organs at risk (OAR)) directly acts as an input for treatment planning. Hence, the need for methods that provide accurate segmentation with minimal user input is obvious. Most frequently used segmentation approaches in radiotherapy are (multi) atlas based segmentation (MABS) strategies and methods based on statistical models of shape and/or appearance (SSM, SAM). MABS offers a high level of robustness and usually does not need any user interaction. Model based approaches on the other hand provide closed surface and anatomically plausible shapes, which is a desirable property for most applications.

Despite their advantageous properties, both approaches suffer from shortcomings when the anatomical structure of interest shows high morphological variability. Most methods for deformable registration have problems to correctly deform structures showing high shape variability. On the other hand the creation of statistical shape models for structures with highly varying shape or even topology is extremely challenging and often requires a large number of labeled training images.

In this paper a novel approach for the fully automatic segmentation of organs at risk in CT images is introduced. Using the proposed method, image patches with different sizes and shapes are extracted for voxel neighborhoods in predefined regions of interest. Using these patches, manifold learning is performed by using Laplacian Eigenmaps in order to obtain the embedding coordinates of each voxel/patch in the patch manifold space. The low-dimensional embedding is then used in combination with locality information to train a Random Forest Classifier in a two-step approach. During the training phase, multiple Random Forests (RF) are trained in order to assess the correct labeling for 9 different locations within a voxel neighborhood. The final class label for each voxel is obtained from the probability maps coming from all 9 random forest classifiers.

The presented approach is evaluated by segmenting the parotid glands in 17 head neck CT scans of cancer patients. Being the major salivary gland, parotid glands are highly critical OARs for radiotherapy of head and neck cancer patients. Unintended irradiation of the parotid glands can result in difficulties for mastication and deglutition. Apart from this, the automatic segmentation of parotid glands is highly challenging. This is due to high anatomical shape variability of the parotid gland on the one hand and low soft-tissue contrast as well as image noise (caused by dental implants) on the other hand.

1.1 Related work

The usage of random forests [1] is highly popular in numerous machine learning applications - not only, but also in medical imaging. Criminisi proposed random forests for the localization of anatomical organs [2], [3]. In addition, random forests were used for segmentation, e.g. for the segmentation of brain tumors [4] or prostates in MRI [5]

Patch based segmentation has also been applied frequently for different anatomical structures [6]–[8]. In [9] the usage of geodesic distances in combination with patch based segmentation and spatial-context information was proposed. However, the usage of this spatial information is dependent on user defined landmarks or additional labeled structures. In [10] the benefit of different features for patch based segmentation has been evaluated for the segmentation of the parotid glands leading to very accurate segmentation results. In addition, patch based segmentation was used in combination with atlas based segmentation [7], [11], [12] as an additional refinement step.

In [13] class probabilities have been used to train a random forest classifier for infant brain segmentation. In [14] iterative, voxel-wise classification was applied for the segmentation of the prostate. However, using this approach segmentation is dependent on previously segmented prostates of the same patient.

A combination of Laplacian Eigenmaps [15] and Random Forests for segmentation has been proposed in [16]. However, the approach is based on computing Laplacian Eigenmaps using whole images instead of patches as input. This does not only result in high computational demands, but also requires a large set of labeled training images.

Summing up, both Random Forest Classification/Regression and Manifold Learning are very active fields of research. However, to the best of our knowledge, the presented combination of 1) Manifold Learning based on multi-scale image patches, 2) Iterative Random Forest Training based on embedding coordinates resulting from Laplacian Eigenmaps and 3) patch based label fusion using multiple random forests 4) without the need for prior elastic registration has not been presented before. The evaluation will show that the proposed method provides accurate segmentation results and performs equally well or better than existing methods for the segmentation of parotid glands in CT images.

2 Methods

2.1 Preprocessing and patch extraction

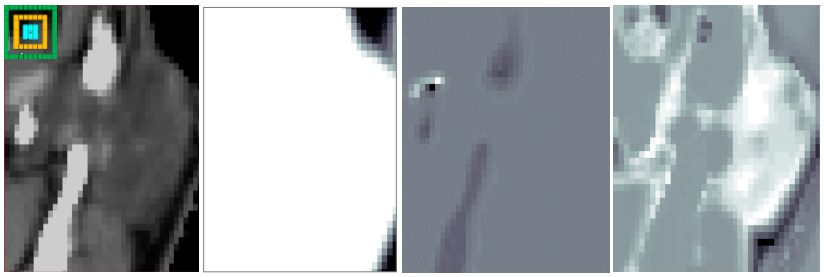
In a first preprocessing step, all images of the training set are registered onto a reference dataset using a rigid transform. In order to account for image noise Mutual Information (MI) is used as a metric. Registration is performed separately for left and right parotid gland using different sub-regions (defined in the reference dataset) for MI computation.

After registration, patches are extracted for each voxel within a predefined region of interest (ROI) containing the respective structure(s) of interest (in this project the parotid gland). The ROI is defined in the reference image and must have the same size for all training and test images. Patches with a 3x3x3 (type A), 7x7x3 (type B) and 11x11x5 (type C) neighborhood are extracted for each voxel. In order to avoid redundancies in overlapping regions and also to reduce the computational burden for the following steps, only patch type A contains all voxels within the patch. Patch types B and C only contain the border voxels of the respective patches (see also fig. 1 (left)). Anisotropic size of the patches is chosen to account for the anisotropic resolution of the images used in this project (1.2x1.2x2.5mm voxel size). Using these patches independently for the following manifold learning step shall on the one hand provide information about small structures (type A) and on the other add knowledge about the appearance of neighboring structures as well as robustness towards local noise (types B and C)

2.2 Manifold learning

It has been shown in various research projects, that non-linear manifold learning approaches like Laplacian Eigenmaps (LapE) [15] can successfully be used for unravelling the inherent structure of the input data [17].
LapE compute a low-dimensional representation of the data in which the distances between a datapoint and its k nearest neighbors are minimized. First a neighborhood graph G is computed, in which every datapoint is connected to its k nearest neighbours. Based on using the Gaussian kernel function the weights for each edge in G are computed and stored in adjacency matrix W. For the computation of the low dimen-

sional representations y_i the cost function $\phi(Y) = \sum_{ij} \|y_i - y_j\|^2 \omega_{ij}$ is used. Using spectral graph theory, the minimization of the cost function can be defined as an eigenproblem so that minimizing $\phi(Y)$ is proportional to minimizing $Y^T L Y$, where the graph Laplacian $L = M - W$. Degree matrix M is a diagonal matrix, whose entries are the row sums of W. By solving $Lv = Mv\lambda$, d eigenvectors $v_i$corresponding to the d smallest eigenvalues can be calculated. As a result, each patch p can be represented by its respective position in patch space. Fig. 1 is presenting the result of manifold learning based on LapE using patch type A as input. Looking at fig. 1, one can see that the resulting eigenvectors have the potential to separate anatomically meaningful structures like skin, bone or different types of soft tissue.



sample *EV1* *EV2* *EV3*

Fig. 1. Illustrating the output of LapE by projecting the embedding coordinates back into image space (sample: a sample from an original image with an illustration of the applied patch configuration in the upper left corner;EV1-EV3: respective embedding coordinates along the coordinate axis formed by 3 exemplary eigenvectors for patch type A)

In this project, LapE are calculated independently for the 3 patch types A, B, C. The resulting embedding coordinates for each patch type are used as feature vector f for the following classification step based on the usage of Random Forests. In this project a value of $d = 20$ was chosen, which results in 60 image features per patch.

In addition to the image features, the offset of each voxel relative to the corner voxel of the predefined ROI is also added to the feature set. By this means, two patches which have similar appearance, but come from very different locations, can be more easily distinguished. Apart from this, the location information can be regarded as a "weak" shape model, which assists in inferring the location (and vague shape) of the parotid gland. Usage of this additional locality information is also the reason why rigid pre-registration was performed during preprocessing.

2.3 Voxel Classification using Random Patch Forests

Random Forest Classification is applied in order to determine a class label $c \in C$ for a given test voxel $x \in \Omega$.

Training. Using a feature representation f for each voxel of the training set, each tree t learns a weak predictor $p(c|f)$. In each node of a tree, the set of training voxels

is split based on a binary test using $s(x; f_k) < t$, where t is a threshold value for each split node and f_k represents the k^{th} feature of f. Based on the test result, the respective voxel is sent to one of its child nodes. During training f_k and t are optimized in a way that the information gain among class distributions is maximized [16]. For improved generalization ability, randomness is induced by only using a randomly sampled subset of f to optimize the parameters for each split node. Optimization is stopped when the maximum information gain falls under a predefined threshold. Growth of a tree stops when a predefined depth is reached or when a node contains less than a predefined number of training samples.

Testing. During testing each new voxel x of an unseen image is pushed through each trained tree by applying the learnt split parameters f_k and t. For each tree the tested voxel will read different leaf nodes, which are associated with different class probabilities. The final probability for a test voxel based on a single random forest is calculated using $p_F(c|x) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|f)$, where T is the number of trees that are used.

Multiple Forests. In the presented approach, classification for each voxel is obtained by using multiple forests. Instead of training only one forest to obtain the label for the center voxel of an image patch, 8 additional RF are calculated in order to assess the class label for the corner voxels of a 3x3x3 neighborhood. Using highly overlapping patches, the probability for a given voxel can then be obtained by $P_m(c|x) = \frac{1}{9}\sum_{t=1}^{9} p_F(c|f)$. The usage of multiple forests is supposed to increase the robustness of the final labeling especially in the presence of image noise compared to using only one random forest.

Two Step Random Forest training and testing. Similar to the approach presented in [13], a two-step approach is used for RF training and testing: In a first training iteration a feature vector $f = \{p_1^A, \dots, p_{20}^A, p_1^B, \dots, p_{20}^B, p_1^C, \dots, p_{20}^C, i_x, i_y, i_z\}$ is used. p_d^L refers to the features obtained for each patch type using Laplacian Eigenmaps and $\{i_x, i_y, i_z\}$ refer to the offset of each voxel relative to the center (voxel) of the predefined ROI (see also section 2.2).Using the Multiple Random Forest approach described above 9 probability maps $P_{f\,1..9}^1$ for each training dataset can be obtained.

For the second training iteration the probability maps $P_{f\,1..9}^1$ are also fed into the classifier in addition to the features used in the first iteration. Based on the fact that similar shaped structures should also result in similar probability maps, the inclusion of the probability maps in the second RF training adds additional locality and shape information to the feature vector, which is expected to be useful for improved RF classification.

During testing, the same two step approach that was used for RF training is applied in order to obtain a final probability map P_m and a final class label C for each x in an unseen image. C is obtained by using

$$C\big(p(x,f)\big) = \begin{cases} 1, p(x,f) > 0.5 \\ 0, p(x,f) \leq 0.5 \end{cases} \quad (1)$$

3 Results

Evaluation scenarios: On the whole, 4 different scenarios have been evaluated in order to test the influence of the suggested contributions:

- **I)** Only LapE embedding coordinates for patch type A in combination with 1 training iteration and 1 RF (for center voxel) are used
- **II)** Equal to I, but using LapE embedding for all patch types
- **III)** Equal to II, but using 9 RF
- **IV)** Equal to III, but using 2 RF iterations.

During evaluation only one LapE had to be computed due to the usage of a leave-one-out strategy. In other settings out-of-sample extension for LapE could be applied for improved efficiency.

Data and parameter settings. The presented approach has been evaluated using 17 CT images of the head neck region with manually labeled parotid glands that acted as ground truth. Image resolution is 1.25x1.25x2.5mm. Accuracy of the segmentation is measured using the DICE coefficient (=$2|A \cap B|/(|A| + |B|)$, where A represents the ground truth and B the results using the proposed approach. In addition, maximum Hausdorff Distance is used to quantify the largest segmentation error.

Pre-registration of the test image to a randomly chosen training image has been performed using rigid transform in combination with Mutual Information metric (to account for the significant noise level in some of the test images). The size for the ROI of which the patches have been extracted was 50x50x87.5mm left Parotid) and 68x75x87.5mm (right parotid) respectively. Based on observations in preliminary test runs the size of the neighborhood for the computation of the Laplacian Eigenmaps was set to $n = 50$. As already mentioned, 20 embedding coordinates have been calculated for each patch type. A leave-one-out cross validation approach was used for the evaluation of all 4 scenarios.

Figure 2 presents results for all 4 testing scenarios. More specifically, the average Dice coefficient and average values for max. Hausdorff distance for left and right parotid glands are depicted. Error bars indicate 25th and 75th percentile.

Figure 3 shows the final labels as well as the final probability maps (for one sample region of an exemplary dataset) resulting from RF classification using scenarios II-IV.

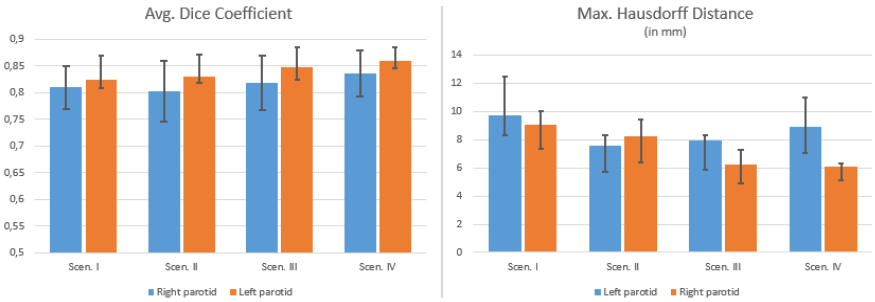


Fig. 2. Comparing average Dice scores and max. Hausdorff distances obtained for left and right parotid glands for all 4 scenarios explained in section 3.

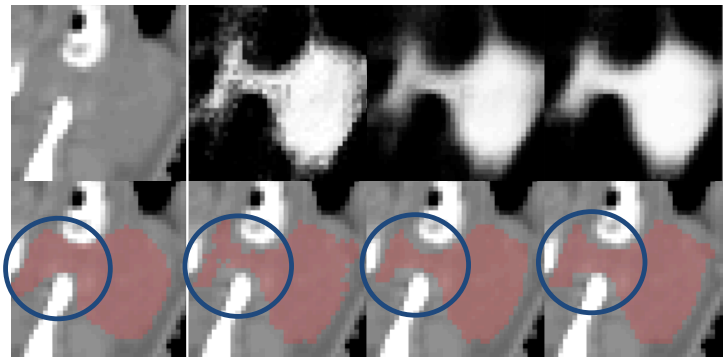


Fig. 3. First **col.**: Sample of original image (top row) and manual segmentation (bottom row); **second to fourth col.**: final probability maps (top row) and corresponding segmentation results (bottom row) obtained with scenarios II-IV, blue circles mark an exemplary area in which gradual improvements of the segmentation accuracy from scen. II – IV can be observed best

4 Discussion

A novel approach based on the usage of manifold learning in combination with a multi-step learning approach based on random patch forests has been presented. The evaluation has shown that the presented method is highly suitable for the segmentation of the parotid glands in CT scans. It could be shown, that the different extensions based on using of multi scale patches (scenario II), an ensemble of random forests (scenario III), as well as the usage of a two-step Random Forest classification scheme (scenario IV) leads to an increase of the respective Dice scores. It can also be observed, however, that max. Hausdorff Distances are lowest using Scenario II for the right parotid gland. This might be due to a smoothing effect that occurs in scenarios III and IV. This smoothing effect (which is also visible in fig. 3) potentially caused that single voxels which are labeled as foreground in scenario II are labeled as back-

ground in scenario III. That has very little or no effect on real segmentation quality, but a positive effect on max. Hausdorff values. This is also supported by the fact that Dice scores are higher using scenarios III and IV. Looking at fig. 3, it can also be observed that the two-step RF classification used in scen. IV results in more distinct probability maps leading to higher segmentation accuracy compared to the ground truth.

Compared to other approaches for parotid gland segmentation which used the same dataset [18] or a (sub)-set of the dataset used in this work [7] for evaluation the presented approach provides equal or higher Dice scores and smaller Hausdorff distances. Also compared to the Dice scores presented in other publications [18]–[20] the obtained results are highly competitive. However, it is also evident that several aspects of the presented approach can potentially be improved: e.g. alternative patch configurations will be tested in order to further increase the positive impact of using multi-scale patches. Moreover, the usage of additional locality information and more sophisticated voting schemes for classification based on multiple RFs as well as the application of different sampling approaches for increased efficiency will be evaluated.

References

1. L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
2. O. Pauly, B. Glocker, A. Criminisi, D. Mateus, A. M. Möller, S. Nekolla, and N. Navab, "Fast multiple organ detection and localization in whole-body MR dixon sequences," in *LNCS,* 2011, vol. 6893 LNCS, pp. 239–247.
3. A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Med. Image Anal.*, vol. 17, pp. 1293–1303, 2013.
4. D. Zikic, B. Glocker, and A. Criminisi, "Atlas encoding by randomized forests for efficient label propagation," in *LNCS,* 2013, vol. 8151 LNCS, pp. 66–73.
5. S. Ghose, J. Mitra, A. Oliver, R. Marti, X. Llad, and I. Segmentation, "A random forest based classification approach to prostate segmentation in MRI," *MICCIA Grand Challenge*, 2012.
6. Z. Wang, C. Donoghue, and D. Rueckert, "Patch-based segmentation without registration: Application to knee MRI," in *LNCS*, 2013, vol. 8184 LNCS, pp. 98–105.
7. C. Wachinger, G. C. Sharp, and P. Golland, "Contour-driven regression for label inference in atlas-based segmentation," in *LNCS*, 2013, vol. 8151 LNCS, pp. 211–218.
8. P. Coupé, J. V. Manjón, 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, pp. 940–954, 2011.
9. Z. Wang, K. Bhatia, B. Glocker, and A. Marvao, "Geodesic patch-based segmentation," MICCAI 2014 (pp. 666-673) ,2014.
10. C. Wachinger, M. Brennan, G. Sharp, and P. Golland, "On the Importance of Location and Features for the Patch-Based Segmentation of Parotid Glands," *people.csail.mit.edu*.
11. X. Han, M. Learning, and in M. Imaging, "Learning-boosted label fusion for multi-atlas auto-segmentation," *Mach. Learn. Med. Imaging*, 2013.
12. 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," *IEEE Trans. Med. Imaging*, vol. 32, pp. 1302–1315, 2013.
13. L. Wang, Y. Gao, F. Shi, G. Li, J. Gilmore, W. Lin, D. Shen, "LINKS:learning-based multi-source integration framework for segmentation of infant brain images," *Neuroimage*, 2015.
14. Y. Gao, S. Liao, D. Shen, and M. physics, "Prostate segmentation by sparse representation based classification," Medical physics, 39(10), 6372-6387.
15. M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, vol. 15. pp. 1373–1396, 2003.
16. H. Lombaert, D. Zikic, A. Criminisi, and N. Ayache, "Laplacian Forests: Semantic Image Segmentation by Guided Bagging," MICCAI 2014 (pp. 496-504) 2014.
17. P. Aljabar, R. Wolz, and D. Rueckert, "Manifold learning for medical image registration, segmentation, and classification," Medical Imaging Intelligence and Analysis, 351.
18. K. Fritscher, M. Peroni, P. Zaffino, M. Spadea, G. C. Sharp, R. Schubert, "Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours ," *Medical Physics, 41,* 2014.
19. V. Fortunati, R.F. Verhaart, F. van der Lijn, W.J. Niessen, J.F. Veenland, M. M. Paulides, T. van Walsum, "Tissue segmentation of head and neck CT images for treatment planning: A multiatlas approach combined with intensity modeling," *Medical physics* 40.7 (2013)
20. A. A. Qazi, V. Pekar, J. Kim, J. Xie, S. L. Breen, and D. A. Jaffray, "Auto-segmentation of normal and target structures in head and neck CT images: a feature-driven model-based approach.," *Medical Physics*, vol. 38, no. 11, pp. 6160–70, Nov. 2011.