A novel atlas-selection approach for multiple atlas segmentation based on Manifold Learning and Random Forests using Multi-Scale Image Patches

P. Raudaschl[1], K. Fritscher[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. Atlas-based segmentation is a frequently used approach in medical imaging and multi atlas-based segmentation (MABS) has achieved great success for various applications. In order to simultaneously exploit the capabilities of MABS, limit execution time and maintain robustness, it is preferable to select a (preferably small) subset of atlases to be used for segmentation.
In this work, an atlas selection strategy using Manifold Learning and Random Forest Regression is presented. The approach aims at learning the relationship between the pairwise appearance of structures and the Dice coefficients of their respective labeling. For this purpose, multi-scale patches in relevant regions of interest are extracted. Local patch models based on linear and non-linear dimensionality reduction are created. Resulting coordinates of the patch embeddings are used to train a regression model for Dice score prediction using Random Forests. Predicted Dice scores are used to rank/select atlases for MABS.
The newly developed approach is applied for segmentation of the left and right parotid glands in CT images of cancer patients. Quantitative evaluation shows that the presented atlas selection approach performs distinctly better than other commonly used selection strategies, especially for a small number of atlases.

Keywords: Atlas-based segmentation, Atlas selection, PCA, Machine Learning

1 Introduction

Image segmentation is a frequently applied task in medical imaging, especially in radiotherapy. Automated segmentation is a substantial component of image-guided adaptive radiotherapy (e.g. for segmentation of organs at risk). The utilization of a priori knowledge of structures that should be segmented is the basis to accomplish a reliable and robust automated segmentation result. In case of atlas-based segmentation this knowledge is available through already segmented atlas images. The segmentation of structures in new images is performed by registering these new images to an already segmented atlas image. Especially for subjects with high inter-structure variability, multi atlas-based segmentation (MABS) approaches have shown to be more accurate than single atlas-based segmentation attempts [1, 2]. Apart from the selected

approach for elastic registration, the selection of an appropriate voting strategy and the selection of atlases have an important influence on the segmentation accuracy and reliability. Moreover, restricting registration to an appropriate subset of atlases can significantly improve segmentation speed. The presented approach is used for segmenting parotid glands in head and neck CT scans of cancer patients. Being the largest salivary glands, the parotid glands are a highly critical OAR. Moreover, due to the high anatomical shape variability of parotid glands, low contrast and image noise caused by dental artifacts, automated segmentation of the parotids is very challenging.

1.1 Related Work

Common atlas selection approaches use image similarity metrics for atlas selection. Typical image similarity metrics are sum of squared distances (SSD), cross correlation [3], and normalized mutual information (NMI) [4]. Based on similarity values, a ranking can be built and used for atlas selection. In a recent work, an atlas selection strategy based on the correlation of inter-atlas similarities was proposed [5]. This approach outperformed common NMI-based atlas selection, especially if only a smaller number of atlases was used. Wolz et al. developed an atlas selection approach based on learning an embedding for atlas selection [6]. Their results have shown an increasing gain in accuracy with increasing distance between the new image that should be segmented and the atlas images compared to common multi-atlas segmentation approaches. In a work by Cao et al. images are projected into a low-dimensional manifold [7]. They used Euclidian distance in this low-dimensional space to determine intrinsic image similarity that is used as an atlas selection criterion. Sanroma et al. [2] used a learning based approach to rank atlases. They tried to learn the relationship between the pairwise appearance of analyzed images and the final labeling performance expressed by Dice values.

1.2 Contribution

In this work an atlas selection strategy using Manifold Learning and Random Forest Regression is presented. The Dice score has shown to be a good (virtual) parameter for atlas selection [2, 5]. Hence, in this work the Dice score was chosen as target parameter for prediction through Random Forest Regression. In contrast to the inspiring approach of Sanroma et al. [2], where HOG features are extracted out of the images and used for model learning with Support Vector Machines, in this work features are generated by using Manifold Learning in combination with multi-scale image patches and Random Forest Regression. Multi-scale image patches are used to assess (semi-) local image characteristics. The patches form the input for a Manifold Learning step. In contrast to [2], the signed distances between the embedding coordinates of the ensemble of all patches extracted from a pair of images will be used to learn (training) and predict (testing) their respective Dice scores. The predicted Dice scores are used to rank and select atlases for MABS. To the best of our knowledge there is no atlas selection approach that uses Manifold Learning with multi-scale image patches in combination with PCA or Laplacian Eigenmaps and Random Forests.

2 Methods

In Fig. 1 the principle overview of the developed atlas selection approach can be seen.

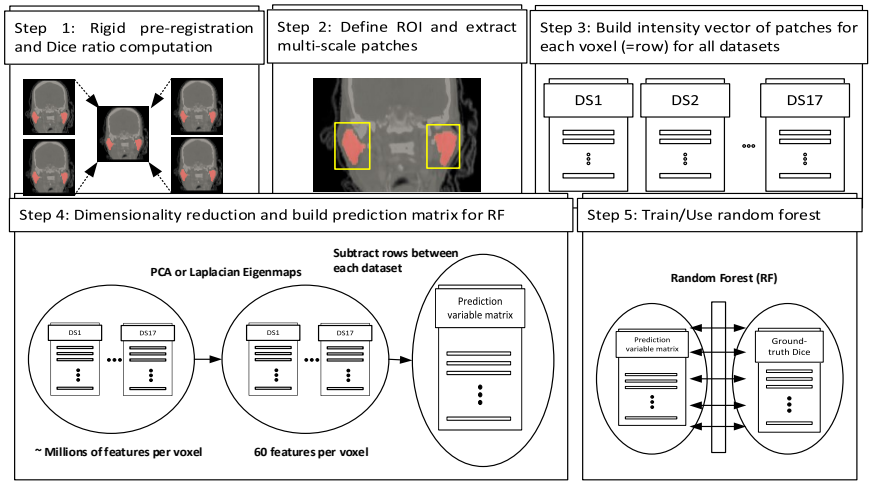

Fig. 1. Principal overview of the proposed atlas selection approach for MABS.

2.1 Data Preprocessing

In the first step all atlas images are rigidly registered on a randomly selected reference atlas image using MI. In order to improve image registration quality, metric computation was restricted to a masked region around left and right parotid gland. Subsequently, regions of interest (ROI) containing all voxels for which patches shall be extracted are defined. The size of the ROIs was constrained to be equal for atlases and target images. Dice ratio, which will be used for atlas ranking (section 2.3) can be calculated as

$$Dice = \frac{2|A \cap B|}{|A| + |B|}. \quad (1)$$

A and B are labeled regions that are compared.

2.2 Multi-scale Patch Extraction

The second step is the extraction of three different image patches for each voxel within the predefined ROI (see Fig. 2). The smallest patch has a cubic structure (patch 1: 3x3x3 voxels). The two remaining patches have a cuboid-like structure: However, only margin voxels that form a hull around the cuboid are used for the larger patch types. These patches are surrounding patch type 1 and have a size of 7x7x3 (patch2)

and 11x11x5 voxels (patch 3). In order to take account of anisotropic image resolution, patch sizes are also anisotropic.

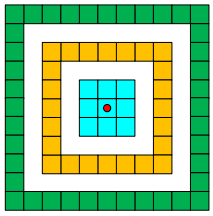

Fig. 2. Principle 2D structure of multi-scale image patches. Red dot: current investigated voxel. Cyan: patch 1. Orange: patch 2. Green: patch 3. Thickness of patches 2 and 3 is one voxel.

By using the border voxels of the cuboid-like patches instead of cubic patches a multi-scale approach can be applied without introducing redundancy compared to using only cubic patches (due to overlapping regions). Furthermore, the number of variables for subsequent dimensionality reduction and Random Forest learning is distinctly lower. This has a significant positive effect on computation time. For each voxel within the predefined ROI the intensity values of its respective patches are arranged in a feature vector $\boldsymbol{t_i}$ for all p atlas datasets for each patch type i. The length of these vectors $\boldsymbol{t_i}$ is m, where m = number of patches per training image x number of voxels per patch type.

2.3 Patch Model Creation

In order to create a patch model, two different approaches have been evaluated. On the one hand, Principal Component Analysis (PCA) [8] for performing linear dimensionality reduction and on the other hand Laplacian Eigenmaps (LapE) [9, 10] providing non-linear embeddings of the input patches. For both approaches, all p feature vectors $\boldsymbol{t_i}$ of patch type i are arranged in three matrices M_i, which are used as input for PCA and LapE.

Principal Component Analysis. PCA aims at identifying the linear combinations of the original variables which maximize their variance. The distribution of data is estimated by the mean value $\boldsymbol{x} = \varepsilon(\boldsymbol{x}) = (\varepsilon(x_1)\varepsilon(x_2) \ldots \varepsilon(x_m))^T \in \mathbb{R}$ and the covariance of the data between the vectors $c_{i,j} = \varepsilon((x_i - m_i)(x_j - m_j))$. Covariances can be represented by the covariance matrix $\boldsymbol{C} = \varepsilon((\boldsymbol{x} - \boldsymbol{m})(\boldsymbol{x} - \boldsymbol{m})^T) \in \mathbb{R}^{mxm}$. From the covariance matrix the eigenvectors and eigenvalues can be calculated. Finally, only the d eigenvectors are chosen that cover a certain (high) degree of all variations. In this project a value of d=20 was chosen.

Manifold Learning. Laplacian Eigenmaps (LapE) are used for non-linear Manifold Learning [9, 10]. A low-dimensional representation of the data is calculated in which the distances between a data point (=patch) and its k nearest neighbors are minimized. LapE generate a graph G based on the neighborhood information. Each data point is represented by a node. The connectivity is determined by a k-nearest neighbor search. Connectivity is weighted by a Gaussian kernel and is stored in the adjacency

matrix $\boldsymbol{W}$. For dimensionality reduction the solution of the minimization problem of the cost function $arg\ min_Y \sum_i \sum_j ||\boldsymbol{y}_i - \boldsymbol{y}_j||^2 w_{i,j}$ can be computed by the eigensystem $\boldsymbol{Lf} = \boldsymbol{\lambda f}$. $\boldsymbol{L}$ is the non-normalized graph Laplacian defined by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$. $\boldsymbol{D}$ is the diagonal matrix with $d_{ii} = \sum_i w_{j,i}$. By solving the generalized eigenvalue problem $\boldsymbol{Lf} = \boldsymbol{\lambda DF}$ d eigenvectors f_i corresponding to the d smallest eigenvalues can be calculated. Consequently, each patch can be represented in the respective position in the patch space. In this project, the number of neighbors used for LapE computation was set to 50.

Random Forest Regression. The first d=20 principal components (PCs) or eigenvectors (EVs), respectively, are used. Using the first 20 PCs/EVs for each patch type, 60 features (= coordinates in PCA space and embeddings resulting from LapE respectively) per voxel are used for Random Forest Regression. For this purpose they are again arranged in vector form. Since Dice ratios are based on pairwise comparison of voxel quantities, also pairwise differences of features between all pairs of images of the training set are calculated and stored in the prediction matrix $\boldsymbol{P}$.

Random Forests (RF) are an ensemble learning approach that can be used for classification and regression [11]. RFs are composed of a multitude of decision trees, where each tree t consists of nodes. During training, the prediction matrix $\boldsymbol{P}$ (containing the embedding coordinates for each patch resulting from PCA and LapE) is trained for predicting the Dice value of each image pair within the training set. The set of embedded coordinates of the training sets gets binary split into two parts at each node of a tree and the respective input is then assigned to one of its child nodes. The target variable of RF regression is the Dice ratio dr after rigid pre-alignment. j is a split variable out of the feature space and s is a splitting point. A binary splitting plane can then be defined: $R_1(j,s) = \{X | X_j \leq s\}$ and $R_2(j,s) = \{X | X_j > s\}$. The splitting variable j and splitting point s are determined by

$$\min_{j,s}\{\min_{c_1}(\sum_{p_i \in R_1(j,s)}(dr_i - c_1)^2) + \min_{c_2}(\sum_{p_i \in R_2(j,s)}(dr_i - c_2)^2)\}. \qquad (2)$$

In each step an optimal pair (j,s) is searched.

For new target images the features have to be computed using PCA or LapE. During testing, new embedding coordinates for the patches of an unseen image are evaluated by the trained trees. By this means a prediction for the "virtual" Dice ratio between (not yet existing) parotid glands label in the new image and every image in the training set can be computed.

2.4 Atlas Selection

Based on the predicted Dice ratios resulting from RF regression an atlas ranking is built for atlas-based segmentation. Higher predicted Dice ratios are ranked first.

3 Results

3.1 Evaluation Strategy

In order to evaluate the segmentation results of the proposed method the results were compared with the following atlas selection strategies:

- **Atlas selection based on original Dice values:** Dice values after deformable registration and segmentation were used to build an atlas ranking. Indeed, this is an unrealistic approach, however, for testing purposes this is often used as a reference. In the following this selection strategy will be referred to as "Oracle" selection.
- **Atlas selection based on common image similarity:** Atlas ranking was generated according to NMI values between image regions. NMI-based selection was chosen as a reference since this is one of the most frequently used selection strategies.

Based on the determined rankings, segmentation performance was tested using different numbers of ranked atlases. Previous results of works with the same datasets have shown that using more than 11 atlases deteriorates segmentation results [5]. Hence, the maximum number of atlases that were tested was set to 11.

A leave-one-out strategy was applied in order to evaluate different atlas selection approaches. The Dice ratio and the 95% Hausdorff distance were used to quantitatively evaluate segmentation accuracy for a varying numbers of atlases

In order to test the significance of the improvement of the respective atlas selection strategy with respect to other atlas selection strategies and different numbers of atlases used for selection, a paired t-test with $p \leq 0.05$ was used [2].

3.2 Data

The atlas database consists of 17 CTs of the head and neck. All images have a voxel size of 1.25x1.25 x 2.5 mm and a dimension of 512 x 512 x 89 slices. Manual delineations of left and right parotid glands were used as gold standard to evaluate segmentation accuracy.

The dimension of the bounding box of the left parotid is 40 x 40 x 35 (56,000 voxels) and of right parotid it is 45 x 50 x 35 (78,750 voxels).

3.3 Experiments

In Fig. 3 (first row) the average Dice ratios for the segmentation of the left and right parotid with respect to different numbers of atlases are visualized. Colored bars represent average Dice ratio for different atlas selection strategies. In Fig. 3 (second row) the average 95% Hausdorff distances of the left and right parotid gland with respect to different numbers of atlases are depicted. Black lines indicate 25^{th} and 75^{th} percentile.

Significance tests performed in order to reveal significant accuracy improvement (in terms of higher Dice scores) for every newly added atlas showed, that no significant improvement could be obtained when using more than 5 datasets. The underlying

atlas ranking for this significance tests is based on Oracle selection, which is considered as "virtual" (i.e. practically not available) gold standard for atlas ranking.

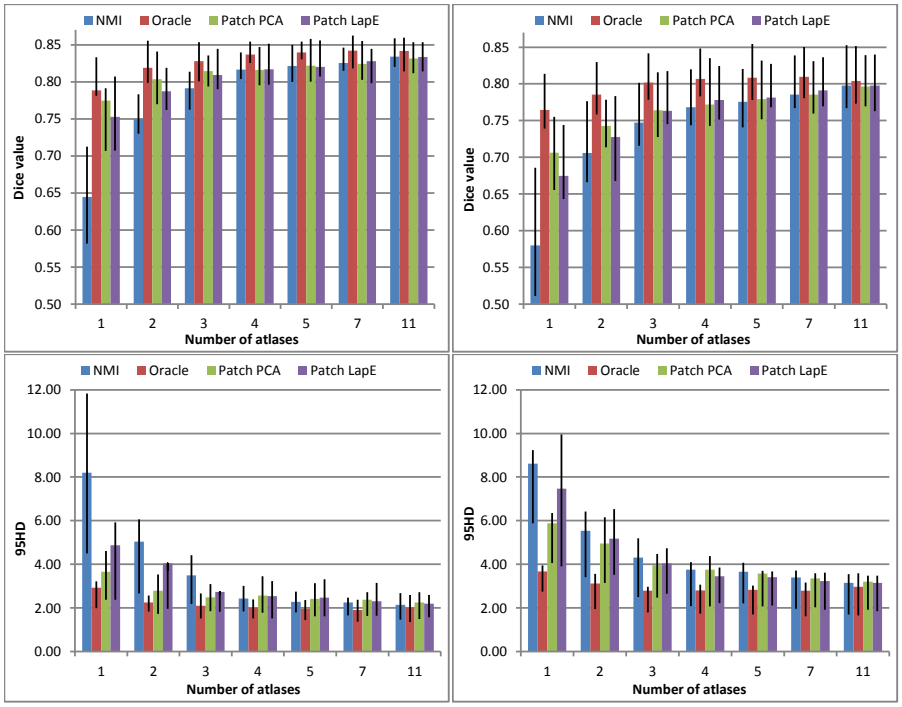

Fig. 3. Evaluation of segmentation performance: First row: Colored bars represent average Dice ratio of the left (first column) and right (second column) parotid gland. Second row: Colored bars represent average 95% Hausdorff distance in mm of the left (first column) and right (second column) parotid. Black lines indicate 25th and 75th percentile.

4 Discussion

Significance tests revealed that using more than 5 atlases does not lead to a significant improvement of the segmentation accuracy even when using Oracle selection. Moreover, evaluation has shown that when using less than 4 atlases the segmentation accuracy is significantly higher when using PCA- and LapE-based atlas selection compared to NMI-based selection (first row in Fig. 3). Furthermore, for the left parotid there was no significant accuracy improvement using 3 atlases based on PCA-based ranking and 5 atlases using NMI-based ranking. The same is true for the right parotid gland when using 4 atlases based on PCA-based ranking. This confirms the positive property of the newly developed approach that equally accurate segmentation results can be achieved with a comparably lower number of atlases.

Comparing PCA vs. LapE of the presented approach, it can be said that PCA-based ranking performs better than the LapE-based ranking when using less than 4 atlases (first row in Fig. 3). For higher number of atlases there are no significant differences concerning segmentation accuracy.

5 Conclusion

It could be shown, that the developed atlas selection technique performs clearly better than NMI-based selection if a low number of atlases is used. In addition, equal segmentation accuracy can be obtained with a lower number of atlases compared to NMI-based selection. Overall, PCA-based atlas selection performed slightly better than LapE-based selection.

In further analyses, the presented approach will be compared to additional selection strategies in addition to NMI. Moreover, more organs in the head neck area will be included in future evaluations. It has to be said, however, that the parotid gland is a very suitable and highly non-trivial test structure, which is frequently used for the evaluation of segmentation approaches for the head and neck region.

6 References

1. van Rikxoort EM, Isgum I, Arzhaeva Y, Staring M, Klein S, Viergever, MA, Pluim, JPW and van Ginneken, B: *Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus.* Medical Image Analysis, Vol. 14, No. 1, pp. 39-49, (2010).
2. Sanroma G, Wu G, Gao Y and Shen D: *Learning to rank atlases for multiple-atlas segmentation.* IEEE Trans Med Imaging, Vol. 33, No. 11, pp. 1939-1953, (2014).
3. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V. and Rueckert, D.: *Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy.* NeuroImage. Vol. 46, No. 3, pp. 726-738, (2009).
4. Studholme C, Hill DLG and Hawkes DJ. *An overlap invariant entropy measure of 3D medical image alignment.* Pattern Recognition. Vol. 32, pp. 71-86, (1999).
5. Raudaschl P, Fritscher K, Zaffino P, Sharp GC, Spadea MF and Schubert R. *A novel atlas-selection approach for multi-atlas based segmentation using the correlation of inter-atlas similarities.* MICCAI - Image-guided adaptive radiotherapy (IGART) workshop, MICCAI 2014, Boston, (2014).
6. Wolz R, Aljabar P, Hajnal JV, Hammers A and Rueckert D. *LEAP: Learning embeddings for atlas propagation.* NeuroImage, Vol. 49, pp. 1316-1325, (2010).
7. Cao Y, Yuan Y Li X, Turkbey, Choyke PL and Yan P. *Segmenting Images by Combining Atlases on Manifold.* MICCAI 2011, pp.272-279, (2011).
8. Rencher AC. Methods of Multivariate Analysis – Principal Component Analysis. John Wiley & Sons, ISBN: 9780471418894, chapter. 12, (2003).
9. Belkin M, Niyogo P. *Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering.* Advances in Neural Information Processing Systems 14, pp. 585-591, (2001).
10. Gerber S, Tasdizen T and Whitaker Ross. *Robust Non-linear Dimensionality Reduction using Successive 1-Dimensional Laplacian Eigenmaps.* Proc. of 24th International Conference on Machine Learning, Corvallis, OR, (2007).
11. Breiman L. *Random Forests.* Machine Learning, Vol. 45, No. 1, pp. 5-32, (2001).