Published in final edited form as:

Phys Med. 2018 August ; 52: 33–41. doi:10.1016/j.ejmp.2018.06.012.

Segmentation of parotid glands from registered CT and MR images

Domen Močnik^{a,*}, Bulat Ibragimov^b, Lei Xing^b, Primož Strojanc, Boštjan Likar^a, Franjo Pernuš^a, and Tomaž Vrtovec^a

^aFaculty of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia

^bDepartment of Radiation Oncology, Stanford University School of Medicine, Stanford, CA, USA

^cInstitute of Oncology Ljubljana, Ljubljana, Slovenia

Abstract

Purpose—To develop an automatic multimodal method for segmentation of parotid glands (PGs) from pre-registered computed tomography (CT) and magnetic resonance (MR) images and compare its results to the results of an existing state-of-the-art algorithm that segments PGs from CT images only.

Methods—Magnetic resonance images of head and neck were registered to the accompanying CT images using two different state-of-the-art registration procedures. The reference domains of registered image pairs were divided on the complementary PG regions and backgrounds according to the manual delineation of PGs on CT images, provided by a physician. Patches of intensity values from both image modalities, centered around randomly sampled voxels from the reference domain, served as positive or negative samples in the training of the convolutional neural network (CNN) classifier. The trained CNN accepted a previously unseen (registered) image pair and classified its voxels according to the resemblance of its patches to the patches used for training. The final segmentation was refined using a graph-cut algorithm, followed by the dilate-erode operations.

Results—Using the same image dataset, segmentation of PGs was performed using the proposed multimodal algorithm and an existing monomodal algorithm, which segments PGs from CT images only. The mean value of the achieved Dice overlapping coefficient for the proposed algorithm was 78.8%, while the corresponding mean value for the monomodal algorithm was 76.5%.

Conclusions—Automatic PG segmentation on the planning CT image can be augmented with the MR image modality, leading to an improved RT planning of head and neck cancer.

*Corresponding author: Domen Močnik, domen.mocnik@fe.uni-lj.si.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest

The authors have no relevant conflicts of interest to disclose.

Keywords

image segmentation; image registration; parotid glands; xerostomia

PACS

87.573nm; 87.57.nj

1. Introduction

Cancer of head and neck (H&N), comprising cancer of pharynx, oral cavity, lips, esophagus, larynx and thyroid, appears in more than one million new cases each year worldwide [1]. Besides surgery and chemotherapy, radiation therapy (RT) is an important treatment of H&N cancer, either as a stand-alone treatment modality or as an adjuvant treatment after surgery. It is estimated that RT is prescribed to approximately 75% of all patients with H&N cancer [2]. Recent external proton beam delivery techniques in RT, such as the *scattering technique* and better yet the *pencil beam scanning*, allow for precise delivering of the radiation energy due to the narrow Bragg peak and creation of a complex spatial radiation dose distribution, which high values highly conform to the tumor shape and sharply decline at the edges of the tumor [3]. The utilization of these advantages requires precise knowledge of the location and shape of the tumor as well as of nearby organs at risk (OAR) for possible dose escalation at the tumor site and reduced induction of toxicity to nearby OARs. This spatial knowledge is obtained from the acquisition of three-dimensional (3D) images of the patient with spatial anatomical or functional information, on which important tumor volumes and OARs are delineated. Based on this information, inverse planning of the optimal dose distribution is performed. Precise delineation of target volumes is thus an important task, as sharp dose gradients can cause parts of tumor underdosage or parts of OAR overdosage. However, manual image delineation is subjective [4–6] and time-consuming [7], therefore reliable and accurate automatic segmentation methods are strongly desired.

Devices for RT planning require delineation of target volumes on 3D computed tomography (CT) image. The CT image contains electron density information and has high contrast for bony structures. In some cases, additional 3D magnetic resonance (MR) image is available, which has higher contrast of soft tissues. It has been demonstrated, for example, that delineation of parotid glands (PGs) and the spinal cord can be easier determined on MR image, which is reflected in better agreement among different observers [4]. There is even an attempt to simulate a planning CT image from MR image in the case of human brain, where physicians are confident that information from MR image alone is sufficient [8]. However, in the case where both image modalities are available, it is wise to incorporate them both in the segmentation procedure, since MR images suffer from mild geometric distortions [9]. To exploit the complementary information of both image modalities and to eradicate the shortcomings of each, the MR image needs to be geometrically aligned (registered) to the CT image [10, 11].

About 40% of all patients treated with RT in H&N suffer from xerostomia (dry mouth), which makes it the most prevalent late complication [12]. It is caused by the radiation damage done to the salivary glands, which impairs saliva secretion from the glands and further leads to indirect complications, such as sore throat, altered taste, dental decay, change in voice quality and impaired chewing and swallowing function. In the stimulated state, during chewing, PGs produce the majority (approximately 60%) of the total saliva [13]. Several studies have shown that a reduction of the mean radiation dose to PGs considerably reduces the incidence and severity of xerostomia, [14–16] and even leads to a recovery of saliva secretion after the initial impairment [12]. It has been further shown that particular sparing of stem and progenitor cells, which reside in the region of the PG containing the major ducts, ensures that the gland continues to produce saliva after RT [17]. Accurate PG segmentation is therefore a crucial prerequisite for avoiding radiation induced xerostomia.

Regarding OARs in H&N, PGs are among the most challenging structures to segment from CT images because of their poor boundary distinction from surrounding tissues and because of their great variation in position, size and shape among different patients [18]. Considerable work has been therefore done to improve the accuracy of PG segmentation. Because RT planning was primarily based on CT imaging, the majority of research, which emphasizes segmentation of OARs in H&N, including PGs, has been focused on segmentation from CT images only [19–28]. However, being soft tissues, PGs have better visibility in MR images, therefore segmentation of PGs from MR images has also been considered [29]. To the best of our knowledge, there has not yet been proposed a method for simultaneous segmentation of PGs from CT and MR images, and therefore we are willing to fill this void by providing a novel approach in this paper.

Magnetic resonance imaging as well as other modalities, such as positron emission tomography (PET), are becoming more accessible and widespread, therefore physicians increasingly utilize them to gather more detailed information about the complex anatomy of PGs in the process of delineation for treatment planning. Since current automatic PG segmentation methods are limited to a single image modality, physicians still vastly rely on manual inspection from multiple modalities to obtain reliable segmentation results. Automatic segmentation processes exploiting multiple modalities have the potential to improve the accuracy and reduce the amount of manual labor, but have to overcome the problem of how to fuse the information from two (or more) different image modalities. A possible solution is to spatially align (register) images of different modalities and then use equally located image intensity-derived features from all modalities in a classification process. This approach was previously used for successful automatic segmentation of H&N tumors from pre-aligned CT, MR and PET images with multichannel Gaussian mixture model [30], and for segmentation of lung tumors from prealigned CT and PET images with various machine learning methods [31]. In recent years, convolutional neural networks (CNNs), a special case of deep learning methods, have demonstrated superior results in image segmentation [32]. Therefore we developed and evaluated a multimodal PG segmentation method that first registers an MR image to a CT image and then uses a CNN to label each voxel from a common image domain according to the resemblance of its

neighborhood appearance in both image modalities to that of voxels from other registered image pairs with known, manual labeling.

The majority of existing PG segmentation methods require registration of many instances from an atlas collection [21–25, 28, 29], which, besides being a relatively slow process, is unsuitable for the development of multimodal segmentation methods, because it may require even more instances (of different modalities) to be registered. In the segmentation phase, our method needs to register only images of the same patient, and in the case of two modalities, this means only one registration. The use of CNNs for multimodal segmentation also provides simultaneous integration of different modalities instead of segmenting individual modality separately and then fusing the resulting segmentations.

2. Materials and methods

2.1. Image registration

To exploit complementary information of both image modalities, the CT and MR image need to be brought into a spatial correspondence in the common image domain. This is done with image registration, a process which seeks the geometric mapping $g: \Omega_R \rightarrow \Omega_M$ from domain Ω_R of the *reference* CT image $I_R: \Omega_R \rightarrow \mathbb{R}$ into domain Ω_M of the *moving* MR image $I_M: \Omega_M \rightarrow \mathbb{R}$, which among all admissible mappings minimizes the criterion function, comprising the similarity term $S(I_R, I_M \circ g)$ and regularization term $R(g)$. The intensities of registered images I_R and $I_M \circ g$ then serve for the segmentation of the reference domain Ω_R .

For the image registration part of our algorithm we use two state-of-the-art algorithms, implemented within the publicly available toolboxes *Elastix* [33] and *ANTs* [34]. *Elastix* was used in the study of multimodal H&N image registration [35] and for most common OARs, including PGs, resulted in a better accuracy than reported in previous studies. *ANTs* is used because of its exceptional performance in the independent evaluation of 14 nonlinear deformation algorithms applied to registration of human brain MR images [36] as well as in the EMPIRE10 challenge [37] for registration of inpatient thoracic CT image pairs. For each available CT and MR image pair we use both toolboxes independently to obtain two different registration results, which are then independently used in the segmentation process.

2.1.1. Registration using Elastix—For the deformation model we choose the initial rigid deformation, followed by the non-rigid free-form deformation comprising tensor products of cubic B-splines [38]. The advanced Mattes mutual information [39, 40] is used as the similarity measure S and optimized using the adaptive stochastic gradient descent method [41]. Both rigid and deformable registration are performed in multiple resolution levels with Gaussian smoothing. Fortunati et al, [35] used the same settings and they also analyzed the impact of the most important parameters on the final registration result. We therefore do not deviate considerably from their choice of optimal parameter values. Gaussian smoothing of both the reference and moving image is performed in three and two resolution levels for the rigid and non-rigid registration, respectively. The B-spline uniform control-point grid spacing is set to 60 mm. The remaining parameters and their values are listed in table 1.

2.1.2. Registration using ANTs—Deformation is composed of the initial rigid deformation, followed by the non-rigid diffeomorphism, obtained by the diffeomorphic symmetric normalization algorithm (SyN) [42]. For both rigid and non-rigid registration, the Mattes mutual information is used as the similarity measure $\mathcal{S}$. The SyN algorithm allows to choose the parameter values related to the linear operator acting on the velocity and/or deformation field, and we do not deviate from the default values here: the variance of the Gaussian filter used for smoothing the velocity field is set to 3 voxel units, while the total field smoothing is not imposed. Both the rigid and diffeomorphic registration are performed in four resolution levels using the Gaussian multiresolution pyramid. The optimization is performed with the gradient descent algorithm. The parameters used for tuning the registration are listed in table 2, together with their values. For their selection we followed the settings used in the EMPIRE10 challenge [43] and recommendations found in works [34, 44] and on ANTs web page [45].

2.2. Segmentation

Let V denote a regular 3D (reference) grid of domain Ω_R , where elements (cells) of $v \in V$ are voxels of the resampled reference image I_R and registered image I_M . Both the reference and registered moving image were resampled, so that the size of $v \in V$ is $1 \times 1 \times 1$ mm³, and V is common to both images. The goal of our segmentation method is to determine for each $v \in V \subset \Omega_R$ whether it belongs to a PG or not. The method is designed in two steps. In the first step, a CNN is used to assign each $v \in V$ its probability $p_v \in [0, 1]$ of affiliation to a PG. The final affiliation of each voxel is determined according to its probability and probabilities of neighboring voxels in the second step. The left and right PGs are segmented separately, but the procedure is equal for both.

2.2.1. Convolutional neural network—A CNN is a type of a feed-forward artificial neural network, used to model the real-valued *response function* $z = z(\mathbf{x}; \mathbf{w})$, where $\mathbf{x}$ is a *feature vector* of the observed element and $\mathbf{w}$ is a vector, consisting of numerical components or *weights* of the CNN. Function z is composed of multiple functions, the so called *layers*, that will be explained briefly in the following paragraphs. In our case, the observed elements are voxels $v \in V$ and their feature vectors consist of image intensity values from their neighborhoods, serving for a numerical description of the appearance of voxel neighborhood in both image modalities.

The use of the CNN takes place in two phases, training and classification. The training phase serves for finding the optimal values for weights $\mathbf{w}$ using prior knowledge, which is in our case a *training set* T of registered CT and MR image pairs with known segmentation. Voxels $v_i, i = 1, 2, \dots, N$, with corresponding labels $y_i \in \{0, 1\}$ and feature vectors $\mathbf{x}_i$ are randomly sampled from reference grids of image pairs from T . Label $y_i = 1$ means that v_i belongs to a PG in its image domain and $y_i = 0$ otherwise. The optimal weight vector $\hat{\mathbf{w}}$ is obtained by minimizing the cross-entropy,

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} - \frac{1}{N} \sum_{i=1}^N \log(P(Y = y_i | \mathbf{X} = \mathbf{x}_i; \mathbf{w})), \quad (1)$$

where the distribution of the conditional probability $P(Y = y | \mathbf{X} = \mathbf{x}; \mathbf{w})$ is calculated from the response function with use of the logistic sigmoid function $\sigma(t) = 1/(1 + e^{-t})$,

$$\begin{aligned} P(Y = 1 | \mathbf{X} = \mathbf{x}; \mathbf{w}) &= \sigma(z(\mathbf{x}; \mathbf{w})), \\ P(Y = 0 | \mathbf{X} = \mathbf{x}; \mathbf{w}) &= 1 - \sigma(z(\mathbf{x}; \mathbf{w})). \end{aligned} \quad (2)$$

In practice this means that function z at optimal $\mathbf{w}$ tries to discriminate voxels according to their feature vectors as high as possible between those that belong to a PG (corresponding to high values of z) and those that do not (corresponding to low values of z).

In the classification phase, given a registered image pair not participating in the training phase, the probability p_v of voxel v belonging to a PG is calculated for every voxel v from the image reference grid as $p_v = \sigma(z(\mathbf{x}_v; \hat{\mathbf{w}}))$, where $\mathbf{x}_v$ is the feature vector of v . In this way, values of p_v (and thus $z(\mathbf{x}_v; \hat{\mathbf{w}})$, since σ is a monotonically increasing function) exhibit a similarity of features of voxel v to features of voxels used in the training phase: higher values exhibit higher similarity to features of 1-labeled voxels and lower values exhibit higher similarity to features of 0-labeled voxels.

The feature vector $\mathbf{x}_v$ of voxel $v = (x, y, z) \in V$ consists of values from six two-dimensional (2D) patches of size 27×27 , where each image modality $I \in \{I_R, I_M \odot g\}$ from the registered pair contributes three patches: $[I(x + i, y + j, z); -13 \dots i j \dots 13]$, $[I(x + i, y, z + k); -13 \dots i, k \dots 13]$ and $[I(x, y + j, z + k); -13 \dots j, k \dots 13]$. The patches therefore consist of CT or MR intensity values from rectangular neighborhoods of size $27 \times 27 \text{ mm}^2$ from axial, coronal and sagittal intersecting planes. This choice of features falls under the so called tri-planar approach, for which it was shown that it achieves similar segmentation performance as the conventional 3D patch approach, but is computationally considerably less demanding [46].

The architecture of our CNN is divided in two parts and is displayed in figure 1. In the first part (first dashed area in figure 1), for each 2D patch of an input feature, a series of consecutive convolution and max-pooling layers are run to obtain a row of 50 high-order features. This part is illustrated for a single patch only, however, there are six such identical parts in parallel, for each 2D patch its own. Together they yield six rows, and in the second part (second dashed area in figure 1), these rows are stacked to obtain a combined 2D feature of size 6×50 , which is then processed using a convolution layer and finally a fully connected layer to obtain the final response of the input feature. The whole CNN process was implemented using the *Microsoft Cognitive Toolkit (CNTK)* [47], an open source deep-learning toolbox.

Convolution layer here is actually a common name for a composition of four consecutive layers, each consuming patches from the previous layer as an input and producing new patches as an output, which are then the input for the next layer. The first and the most important layer applies several 2D convolutions with different kernels on input patches. What follows is the batch-based normalization layer (resulting convolution patches are normalized by their mean patch and their standard deviation patch, and they are afterwards multiplied by one learning constant, while a second learning constant is added). Next is the

activation layer (where the rectified linear unit function is applied to all patches element-wise), and finally the dropout layer (where some elements in patches are randomly eliminated). The max-pooling layers, which are illustrated separately on figure 1, output patches half the size of the input patches with the output elements being the maximum within 2×2 non-overlapping tiles.

For an image pair $\tau = (I_R^\tau, I_M^\tau \circ g^\tau) \in T$ with the corresponding reference grid V_τ , we denote with $S_\tau \subset V_\tau$ a set of all 1-labelled voxels, i.e. all voxels which were on I_R^τ manually annotated (by a physician) as those that belong to the left (or right) PG. For each $\tau \in T$ we sample with replacement a multiset of $k = 20,000$ 1-labeled voxels from S_τ (meaning that each voxel from S_τ can be selected more than one time) and a multiset of k 0-labeled voxels from $B_\tau = \{v \in V_\tau \setminus S_\tau; 2 \text{ mm} \leq d(v, S_\tau) \leq 40 \text{ mm}\}$ where $d(v, S_\tau)$ is the Euclidean distance from voxel v to its nearest voxel from S_τ . Features of a total $k|T|$ 1-labeled and $k|T|$ 0-labeled voxels, not necessarily unique, are used for training of an individual CNN ($N = 2k|T|$). This sampling is used to equalize the number of representatives for both labels, since in general sets S_τ and B_τ have different cardinalities.

2.2.2. Classification post-processing—Once the trained CNN assigns the probabilities p_v to every voxel $v \in V$ of the observed image pair not participating in the training process, the final label $y_v \in \{0, 1\}$ $v \in V$, is determined in a series of post-processing procedures. First, a graph-cut algorithm, using a publicly available software FastPD [48], is run to obtain the connected structures consisting of voxels with high probabilities. If $\mathbf{L} = [y_v; v \in V]$ denotes a binary vector of labels, then, with the graph-cut algorithm, we seek the optimal labeling $\hat{\mathbf{L}}$ as

$$\hat{\mathbf{L}} = \underset{\mathbf{L} \in \{0,1\}^{|V|}}{\operatorname{argmin}} \left(\sum_{v \in V} \phi(p_v, y_v) + \sum_{(u,v) \in E} \psi(u, v) \right), \quad (3)$$

where E is the set of all neighboring voxel pairs,

$$\phi(p_v, y_v) = \begin{cases} 1 - p_v & ; y_v = 1 \\ p_v & ; y_v = 0 \end{cases} \text{ and } \quad (4)$$

$$\psi(u, v) = \begin{cases} \exp\left(\frac{-(p_u - p_v)^2}{2}\right) & ; y_u = y_v \\ 0 & ; \text{otherwise} \end{cases} \quad (5)$$

The minimization of the first term in (3) favors the inclusion of voxels with high probabilities, while the minimization of the second term strives to select a connected component of voxels whose probabilities do not differ considerably. The final set of all

voxels with assigned label 1 consists of several connected components, but only the largest one is chosen and then a series of dilate and erode operations is run on it to fill the possible cavities and to smooth its boundary. The obtained set serves as the final segmentation structure. Figure 2 shows an example of a result of applying described postprocessing procedures on the probability maps, obtained from the CNN.

3. Experiments and results

3.1. Image data

We used a collection of planning images from 44 different patients (25 males, 19 females), scheduled for H&N RT. The age of patients ranged from 12 to 93 years (mean 58.8 years, median 62 years). For every patient, a pair of 3D CT and MR images of the H&N region was available. Computed tomography images were acquired with Somatom Definition AS (Siemens Healthineers, Germany) and Discovery ST (GE Healthcare, USA) scanners, using 120 kVp and 75–441 mAs. Magnetic resonance images were T1 weighted and were acquired using Discovery MR750 (GE Healthcare, USA) scanner. Axial plane resolution varied between $0.781 \times 0.781 \text{ mm}^2$ and $1.27 \times 1.27 \text{ mm}^2$ for CT images, and between $0.312 \times 0.312 \text{ mm}^2$ and $0.898 \times 0.898 \text{ mm}^2$ for MR images. Slice thickness varied between 1.25 and 3.27mm for CT images, and between 4 and 6.5 mm for MR images. An experienced physician manually delineated the left and right PG on CT images. Among those, one image lacked left PG delineation and two other images lacked right PG delineation.

3.2. Evaluation

All available CT and MR image pairs (I_R^K, I_M^K) , $\kappa \in K = \{1, 2, \dots, 44\}$, were registered using both described registration procedures to obtain mappings g^K (using Elastix toolbox) and h^K (using ANTs tool box), $\kappa \in K$. The set K was divided in five disjoint subsets K_i , $i = 1, 2, \dots, 5$, each containing 8–9 indices. For the segmentation of the registered image pairs $(I_R^K, I_M^K \circ g^K)$ with indices $\kappa \in K_i$, image pairs $(I_R^\xi, I_M^\xi \circ g^\xi)$ with indices $\xi \in \bigcup_{j \neq i} K_j$ were used for training the CNN. The same procedure was used for the segmentation of image pairs registered with mappings h^K . Therefore, for each K_i , two CNNs were trained for the segmentation of a single PG, and each PG (left and right) was considered separately, which amounted to a total of $5 \cdot 2 \cdot 2 = 20$ trained CNNs. For the purpose of comparison to our multimodal PG segmentation, we also executed the segmentation of PGs from CT images only, using a similar segmentation process [20].

The quality of automatic segmentation method was measured for each image pair by comparing the resulting segmented region P with the manually delineated region S in terms of the *Dice coefficient* overlap measure $d_{\text{Dice}}(P, S) = 2 |P \cap S| / (|P| + |S|)$ and in terms of the *average symmetric surface distance* (ASSD)

$$d_{\text{ASSD}}(P, S) = \frac{1}{|b(P)| + |b(S)|} \left(\sum_{u \in b(P)} \min_{v \in b(S)} d(u, v) + \sum_{v \in b(S)} \min_{u \in b(P)} d(u, v) \right), \quad (6)$$

where $b(S)$ denotes the set of all border voxels of S and $d(u, v)$ is the Euclidean distance between voxels u and v .

3.3. Results

A total of 43 image pairs were used in the analysis of segmentation of the left PG and 42 image pairs in the analysis of segmentation of the right PG, where 40 image pairs were common. The results of segmentation methods are shown in figure 3, while the comparison of multimodal and monomodal segmentation methods is given in table 3. In terms of the Dice coefficient, the Wilcoxon signed-rank test showed a statistically significant improvement of the multimodal method over the monomodal method in both PGs with p -values being $5.1 \cdot 10^{-5}$ (Elastix) and $3.7 \cdot 10^{-5}$ (ANTs) for the left PG, and $9.8 \cdot 10^{-7}$ (Elastix) and $1.0 \cdot 10^{-6}$ (ANTs) for the right PG. There was no statistically significant difference in the accuracy between the two different registration toolboxes used in our multimodal segmentation method ($p = 0.56$ for the left PG and $p = 0.83$ for the right PG). In terms of ASSD, there was no statistically significant improvement of the multimodal method over the monomodal method, with both p -values being above 0.1.

4. Discussion

In comparison to the results of the monomodal segmentation method, which is limited to the segmentation from CT images only, the results of the multimodal method showed an improvement in the Dice coefficient in 80.6% of the cases. The overall average improvement in the Dice coefficient was 2.23 percentage points. Since xerostomia has such a high incidence, even a small improvement in segmentation results may benefit thousands of patients treated with RT in H&N. The results could be further improved in the ease of better MR image resolution in axial direction because of two reasons. First, patches from coronal and sagittal intersecting planes of MR images would have more distinctive spatial information, possibly increasing the recognition ability of the CNN. Second, registration of the MR image to the CT image could be improved, therefore providing more spatially aligned information to the segmentation method. Among 44 available MR images in our database, only six had axial resolution less than 6 mm and none less than 4 mm. It is fairly common in current MR imaging for axial resolution to be 3 mm or even less.

The achieved mean Dice coefficient results from PG segmentation in other methods are reported to be 65% [23], 76% (from CT) and 79% (from MR) [29], 77% [27], 79% [25], 79% Mid 82% [26], 82% [22], 83% [19, 28] and 84% [21, 24], which make our average result of 78.8% and median result of 80.7% comparable. The majority of existing methods use atlas segmentation as an integral part, meaning that a set of prior segmented images, forming the *atlas*, are registered to the current image, and the propagated volumes from atlas images are used as a basis for further segmentation refinement, using active contours or statistical shape and appearance models. Besides the Dice coefficient, which is the most commonly reported measure for the segmentation quality, several studies also report the surface distance [19, 21, 22, 24, 25, 27–29]. However, its definition is inconsistent and therefore a direct comparison with the obtained results is not always adequate. Nevertheless, the obtained ASSD of 1.57 mm is better than or comparable to any reported surface

distance, e.g. average surface distance of 1.83 mm [19], mean contour distance of 1.80 mm [21], mean absolute distance of 1.55 mm [24], and ASSD of 1.66 mm [27].

The results of our multimodal segmentation methods show no clear distinction in the accuracy between two different registration methods used. Visual inspection reveals that both registration methods provide satisfactory registration in the region of PGs in the majority of cases, probably because of the clear distinction against the nearby mandible in both image modalities. Although the registration results were not perfect, the overall segmentation method in most cases still managed to benefit from the inclusion of additional patches from the MR image in the feature vector of each individual voxel. Figure 4 shows a ease with visually good registration in regions around PGs, which resulted in improved segmentation results of the multimodal method in contrast to the monomodal method. Figure 5 shows a ease with visually slightly worse registration around the right PG, but the segmentation was improved nevertheless, suggesting that perfect registration is not needed, although desired. However, if registration is not good enough, then one can not expect for improved segmentation results, as is the ease shown in figure 6.

Among 11 cases (nine common cases), where at least one multimodal segmentation method (using Elastix or ANTs) showed a drop in the Dice coefficient compared to the monomodal method for the left PG, seven cases already provided a good segmentation result with the monomodal method, with the Dice coefficient ranging from 0.802 to 0.866, which is well above the median value of 0.778. Since visual inspection revealed no deficiencies in registration results for these cases, and because the average drop in the Dice coefficient was only -0.0073 (for Elastix) and -0.0085 (for ANTs), we believe that there simply was no space for further improvements with the multimodal method. For the remaining four cases, two of them visually revealed a worse registration result, and two of them were degenerated cases, which also have a poor monomodal segmentation result (Dice coefficient of 0.602 and 0.618). A very similar observation holds for segmentation of the right PG, where among seven cases (six common), four already provided a good segmentation result with the monomodal method (Dice coefficient from 0.822 to 0.861) with the average drop of only 0.006 (both Elastix and ANTs), while two revealed a poor registration result and one ease was degenerated. On the other hand, no statistically significant improvement in ASSD indicates an insufficient ability of the proposed method to fill some protruding portions of the structure with non-negligible surface area but a relatively small volume. Observe, for example, the bottom row of Figure 2, where the bottom-most part of the reference structure was not included in the final segmentation. In the ease of HT. however, volumetric measures, such as Dice, are more indicative for organ preserving in terms of its geometry.

Since the resulting Dice coefficients and ASSD metric depend on the physician's delineation of PGs on CT images, interobserver variability should be taken into account. Therefore it can not be expected that the delineation agreement between the automatic method and the physician would be considerably better than the delineation agreement between different physicians. The agreements of PG delineation on CT images in terms of the Dice coefficient among different physicians have been reported to be on average 85.3% (80.9% – 89.4%) among three different observers [27], and 80% [26] and 90% [25] among five different

observers. In the last case, however, only images not distorted by tumors or artefacts were used. These findings show that segmentation of PGs is very challenging even for a human.

The methodology used in this paper could easily be used to incorporate alternative or additional image modalities as well, by simply replacing or adding additional input patches from other modalities in the CNN. Our work therefore establishes methodological foundation for adding more image modalities. Joint use of multiple complementary modalities has shown to be beneficial in the segmentation of lung tumor with co-registered PET and CT images [49], H&N tumor with PET and MR images [50], brain tumor [51], globus pallidus and stratum [52] with multiple MR sequences, and optical disc and cup with optical coherence tomography and color fundus photography [53]. The above listed multimodal segmentation studies have an important data advantage, that is, different image modalities of the same patient are either already perfectly aligned because of combined PET/CT or PET/MR acquisition machines, or, as is the case of brain imaging, different MR modalities can be easily aligned with simple rigid registration, because the brain is a rigid organ. Registration of CT and MR images of H&N is a much more difficult problem because of substantial anatomical and positional changes between each image acquisition, such as esophagus peristalsis, tongue position, muscle contractions and different neck tilt. In our opinion, this is also the reason why multimodal image segmentation of OARs in H&N has not yet been exploited. We can expect that improving the multimodal image registration will remain an important task in the multimodal segmentation of OARs in H&N, especially since the above listed studies show beneficial results for well aligned images. While other multimodal methods apply segmentation to each modality separately and then fuse individual segmentations into a final one, the main advantage of our method is the deep interweaving of different image modality patches through many deep layers of CNN, therefore establishing a more deep connection between different modalities.

Since combined PET/CT and PET/MR acquisition machines are becoming more common, adding PET modality could be our further step. The proposed segmentation method can be used for the segmentation of other organs as well. We focused on PGs because current automatic segmentation methods still do not provide satisfactory results for them and, at the same time, they are crucial for preventing very common adverse effects in RT of H&N. Additionally, the segmentation of PGs can largely benefit from the MR image modality because of a better visibility and eradication of dental artifacts, therefore making the MR image a strong addition to the CT image, which is still unavoidable in RT planning. We may analyse the segmentation of other OARs in H&N in the future, but MR images with larger fields of view are needed, since our current MR images have a varying field of view and many do not completely cover some structures. Nevertheless, the design of our method is directed more towards the segmentation of organs with complex shape and appearance, while tumors with compact shape may be successfully segmented e.g. using the level set method [54], that does not require a learning process.

5. Conclusion

We proposed a method for segmentation of PGs from a pair of CT and MR images. The method uses implementations of the state-of-the-art registration algorithms from publicly

available toolboxes Elastix or ANTs to register the MR image to the CT image, and then uses CNNs to classify voxels based on their neighboring planar patch features of image intensities from both (registered) images. The proposed method showed improved results compared to a related CNN-based segmentation method which uses CT modality only. We demonstrated that CNNs have the ability to successfully incorporate additional image modality to aid the segmentation in cases where image registration between image modalities offers satisfactory results. There is a certain confidence that segmentation of other organs with better visibility on MR images could benefit from the proposed method as well.

Acknowledgments

Funding

This work was supported by the Slovenian Research Agency (ARRS) [grant numbers P2-0232, J7-6781, J2-7211, J2-7118 and J2-8173], and by the National Institutes of Health, USA [grant numbers 1R01CA176553 and R01E0116777].

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65(2):87–108. DOI: 10.3322/caac.21262 [PubMed: 25651787]
2. Barton MB, Jacob S, Shafiq J, Wong K, Thompson SR, Hanna TP, et al. Estimating the demand for radiotherapy from the evidence: a review of changes from 2003 to 2012. *Radiother Oncol.* 2014; 112(1):140–4. DOI: 10.1016/j.radonc.2014.03.024 [PubMed: 24833561]
3. Schippers JM, Lomax AJ. Emerging technologies in proton therapy. *Acta Oncol.* 2011; 50838(6): 50. doi: 10.3109/0284186X.2011.582513
4. Geets X, Daisne JF, Arcangeli S, Coche E, De Poel M, Duprez T, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI. *Radiother Oncol.* 2005; 77(1):25–31. DOI: 10.1016/j.radonc.2005.04.010 [PubMed: 15919126]
5. Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D variation in delineation of head and neck organs at risk. *Radiat Oncol.* 2012; 7:32. doi: 10.1186/1748-717X-7-32 [PubMed: 22414264]
6. Perna L, Sini C, Cozzarini C, Agnello G, Cattaneo GM, Hysing LB, et al. Deformable registration-based segmentation of the bowel on megavoltage CT during pelvic radiotherapy. *Phys Med.* 2016; 32(7):898–904. DOI: 10.1016/j.ejmp.2016.06.009 [PubMed: 27345259]
7. Harari PM, Song S, Tomé WA. Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer. *Int J Radiat Oncol Biol Phys.* 2010; 77(3):950–8. DOI: 10.1016/j.ijrobp.2009.09.062 [PubMed: 20378266]
8. Aouadi S, Vasic A, Paloor S, Torfeh T, McGarry M, Petric P, et al. Generation of synthetic CT using multi-scale and dual-contrast patches for brain MRI-only external beam radiotherapy. *Phys Med.* 2017; 42(Supplement C):174–84. DOI: 10.1016/j.ejmp.2017.09.132 [PubMed: 29173912]
9. Torfeh T, Hammoud R, Perkins G, McGarry M, Aouadi S, Celik A, et al. Characterization of 3D geometric distortion of magnetic resonance imaging scanners commissioned for radiation therapy planning. *Magn Reson Imaging.* 2016; 34(5):645–53. DOI: 10.1016/j.mri.2016.01.001 [PubMed: 26795695]
10. Thiagarajan A, Caria N, Sehöder H, Iyer NG, Wolden S, Wong RJ, et al. Target volume delineation in oropharyngeal cancer: impact of PET, MRI, and physical examination. *Int J Radiat Oncol Biol Phys.* 2012; 83(1):220–7. DOI: 10.1016/j.ijrobp.2011.05.060 [PubMed: 22035663]
11. Grégoire V, Langendijk JA, Nuyts S. Advances in radiotherapy for head and neck cancer. *J Clin Oncol.* 2015; 33(29):3277–84. DOI: 10.1200/JCO.2015.61.2994 [PubMed: 26351354]
12. Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a

- phase 3 multicentre randomised controlled trial. *Lancet Oncol.* 2011; 12(2):127–36. DOI: 10.1016/S1470-2045(10)70290-4 [PubMed: 21236730]
13. Eisbruch A, Rhodus N, Rosenthal D, Murphy B, Rasch C, Sonis S, et al. How should we measure and report radiotherapy-induced xerostomia? *Semin Radiat Oncol.* 2003; 13(3):226–34. DOI: 10.1016/S1053-4296C03)00033-X [PubMed: 12903012]
 14. Lee TF, Chao PJ, Ting HM, Chang L, Huang YJ, Wu JM, et al. Using multivariate regression model with least absolute shrinkage and selection operator (LASSO) to predict the incidence of xerostomia after intensity-modulated radiotherapy for head and neck cancer. *PLoS One.* 2014; 9(2):e89700.doi: 10.1371/journal.pone.0089700 [PubMed: 24586971]
 15. Yuan L, Wu QJ, Yin FF, Jiang Y, Yoo D, Ge Y. Incorporating single-side sparing in models for predicting parotid dose sparing in head and neck IMRT. *Med Phys.* 2014; 41(2):021728.doi: 10.1118/1.4862075 [PubMed: 24506619]
 16. Tuomikoski L, Kapanen M, Collan J, Keyriläinen J, Saarilahti K, Loimu V, et al. Toward a more patient-specific model of post-radiotherapy saliva secretion for head and neck cancer patients. *Acta Oncol.* 2015; 54(9):1310–6. DOI: 10.3109/0284186X.2015.1067717 [PubMed: 26203929]
 17. van Luijk P, Pringle S, Deasy JO, Moiseenko VV, Faber H, Hovan A, et al. Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer. *Sci Transl Med.* 2015; 7(305):305ral47.doi: 10.1126/scitranslmed.aac4441
 18. Medbery R, Yousem DM, Needham MF, Kligerman MM. Variation in parotid gland size, configuration, and anatomic relations. *Radiother Oncol.* 2000; 54(1):87–9. DOI: 10.1016/S0167-8140(99)00150-4 [PubMed: 10719704]
 19. Wang Z, Wei L, Wang L, Gao Y, Chen W, Shen D. Hierarchical vertex regression-based segmentation of head and neck CT images for radiotherapy planning. *IEEE Trans Image Process.* 2018; 27(2):923–37. DOI: 10.1109/TIP.2017.2768621 [PubMed: 29757737]
 20. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys.* 2017; 44(2):547–57. DOI: 10.1002/mp.12045 [PubMed: 28205307]
 21. Raudasehl PF, Zaffino P, Sharp GC, Spadea MF, Chen A, Dawant BM, et al. Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. *Med Phys.* 2017; 44(5):2020–36. DOI: 10.1002/mp.12197 [PubMed: 28273355]
 22. Waehinger C, Fritseher K, Sharp G, Golland P. Contour-driven atlas-based segmentation. *IEEE Trans Med Imaging.* 2015; 34(12):2492–505. DOI: 10.1109/TMI.2015.2442753 [PubMed: 26068202]
 23. Hoang Due AK, Eminowicz G, Mendes R, Wong SL, McClelland J, Modat M, et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. *Med Phys.* 2015; 42(9):5027–34. DOI: 10.1118/1.4927567 [PubMed: 26328953]
 24. Fritseher KD, Peroni M, Zaifino P, Spadea MF, Schubert R, Sharp G. Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours. *Med Phys.* 2014; 41(5):051910.doi: 10.1118/1.4871623 [PubMed: 24784389]
 25. Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. *Radiat Oncol.* 2014; 9:173.doi: 10.1186/1748-717X-9-173 [PubMed: 25086641]
 26. Mattiucci GC, Boldrini L, Chiloio G, D'Agostino GR, Chiesa S, De Rose F, et al. Automatic delineation for replanning in nasopharynx radiotherapy: What is the agreement among experts to be considered as benchmark? *Acta Oncol.* 2013; 52(7):1417–22. DOI: 10.3109/0284186X.2013.813069 [PubMed: 23957565]
 27. Faggiano E, Fiorino C, Sealeo E, Broggi S, Cattaneo M, Maggiulli E, et al. An automatic contour propagation method to follow parotid gland deformation during head-and-neck cancer tomotherapy. *Phys Med Biol.* 2011; 56(3):775–91. DOI: 10.1088/0031-9155/56/3/015 [PubMed: 21239848]

28. Qazi AA, Pekar V, Kim J, Xie J, Breen SL, Jaffray DA. Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach. *Med Phys.* 2011; 38(11):6160–70. DOI: 10.1118/1.3654160 [PubMed: 22047381]
29. Wardman K, Prestwich RJD, Gooding MJ, Speight RJ. The feasibility of atlas-based automatic segmentation of MRI for H&N radiotherapy planning. *J Appl Clin Med Phys.* 2016; 17(4):146–54. DOI: 10.1120/jacmp.v17i4.6051 [PubMed: 27455480]
30. Yang J, Beadle BM, Garden AS, Schwartz DL, Aristophanous M. A multimodality segmentation framework for automatic target delineation in head and neck radiotherapy. *Med Phys.* 2015; 42(9): 5310–20. DOI: 10.1118/1.4928485 [PubMed: 26328980]
31. Kawata Y, Arimura H, Ikushima K, Jin Z, Morita K, Tokunaga C, et al. Impact of pixel-based machine-learning techniques on automated frameworks for delineation of gross tumor volume regions for stereotactic body radiation therapy. *Phys Med.* 2017; 42:141–9. DOI: 10.1016/j.ejmp. 2017.08.012 [PubMed: 29173908]
32. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017; 42:60–88. DOI: 10.1016/j.media. 2017.07.005 [PubMed: 28778026]
33. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging.* 2010; 29(1):196–205. DOI: 10.1109/TMI. 2009.2035616 [PubMed: 19923044]
34. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage.* 2011; 54(3):2033–44. DOI: 10.1016/j.neuroimage.2010.09.025 [PubMed: 20851191]
35. Fortunati V, Verhaart RF, Angeloni F, van der Lugt A, Niessen WJ, Veenland JF, et al. Feasibility of multimodal deformable registration for head and neck tumor treatment planning. *Int J Radiat Oncol Biol Phys.* 2014; 90(1):85–93. DOI: 10.1016/j.ijrobp.2014.05.027 [PubMed: 25015199]
36. Klein S, Pluim JPW, Staring M, Viergever MA. Adaptive stochastic gradient descent optimisation for image registration. *Int J Comput Vision.* 2009; 81(3):227. doi: 10.1007/s11263-008-0168-y
37. Murphy K, van Ginneken B, Reinhardt JM, Kabus S, Ding K, Deng X, et al. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE Trans Med Imaging.* 2011; 30(11):1901–20. DOI: 10.1109/TMI.2011.2158349 [PubMed: 21632295]
38. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging.* 1999; 18(8): 712–21. DOI: 10.1109/42.796284 [PubMed: 10534053]
39. Thevenaz P, Unser M. Optimization of mutual information for multiresolution image registration. *IEEE Trans Image Process.* 2000; 9(12):2083–99. DOI: 10.1109/83.887976 [PubMed: 18262946]
40. Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W. PET-CT image registration in the chest using free-form deformations. *IEEE Trans Med Imaging.* 2003; 22(1):120–8. DOI: 10.1109/TMI.2003.809072 [PubMed: 12703765]
41. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage.* 2009; 46(3):786–802. DOI: 10.1016/j.neuroimage.2008.12.037 [PubMed: 19195496]
42. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal.* 2008; 12(1):26–41. DOI: 10.1016/j.media.2007.06.004 [PubMed: 17659998]
43. Tustison N, Song G, Gee J, Avants B. Two greedy SyN variants for pulmonary image registration. http://empire10.isi.uu.nl/pdf/article_antsregistrationb-splinesyn.pdf; Online material. Accessed: January 15, 2018
44. Tustison NJ, Avants BB. Explicit B-spline regularization in diffeomorphic image registration. *Front Neuroinform.* 2013; 7doi: 10.3389/fninf.2013.00039
45. Anatomy of an antsRegistration call. <https://github.com/stnava/ANTs/wiki/Anatomy-of-antsRegistration-call>; Online material. Accessed: January 15, 2018
46. Lai M. Deep learning for medical image segmentation. *CoRR.* 2015. abs/1505.02000. URL: <http://arxiv.org/abs/1505.02000>

47. Eversole Yu D, Seltzer A, Yao M, Kuehaiev K, Zhang OY. , et al. An introduction to computational networks and the computational network toolkit. Tech Rep. 2014. URL: <https://www.microsoft.com/en-us/research/publication/an-introduction-to-computational-networks-and-the-computational-network-toolkit/>
48. Komodakis N, Tziritas G. Approximate labeling via graph cuts based on linear programming. IEEE Trans Pattern Anal. 2007; 29(8):1436–53. DOI: 10.1109/TPAMI.2007.1061
49. Ju W, Xiang D, Zhang B, Wang L, Kopriva I, Chen X. Random walk and graph cut for co-segmentation of lung tumor on PET-CT images. IEEE Trans Image Process. 2015; 24(12):5854–67. DOI: 10.1109/TIP.2015.2488902 [PubMed: 26462198]
50. Leibfarth S, Eckert F, Welz S, Siegel C, Schmidt H, Schwenzer N, et al. Automatic delineation of tumor volumes by co-segmentation of combined PET/MR data. Phys Med Biol. 2015; 60(14):5399.doi: 10.1088/0031-9155/60/14/5399 [PubMed: 26111013]
51. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging. 2015; 34(10):1993–2024. DOI: 10.1109/TMI.2014.2377694 [PubMed: 25494501]
52. Visser E, Keuken MC, Douaud G, Gaura V, Bachoud-Levi AC, Remy P, et al. Automatic segmentation of the striatum and globus pallidus using MIST: multimodal Image Segmentation Tool. Neuroimage. 2016; 125(Supplement C):479–97. DOI: 10.1016/j.neuroimage.2015.10.013 [PubMed: 26477650]
53. Miri MS, Abràmoff MD, Lee K, Niemeijer M, Wang JK, Kwon YH, et al. Multimodal segmentation of optic disc and cup from SD-OCT and color fundus photographs using a machine-learning graph-based approach. IEEE Trans Med Imaging. 2015; 34(9):1854–66. DOI: 10.1109/TMI.2015.2412881 [PubMed: 25781623]
54. Gui L, Li C, Yang X. Medical image segmentation based on level set and isoperimetric constraint. Phys Med. 2017; 42:162–73. DOI: 10.1016/j.ejmp.2017.09.123 [PubMed: 29173911]

Figure 1.
The architecture of the applied convolutional neural network.

Figure 2.

An example of the segmentation of the left parotid gland (PG). The top row shows the probability maps in grayscale encoding, where a brighter intensity at a voxel v corresponds to a higher probability p_v that v belongs to a PG. From left to right, axial, coronal and sagittal cross-sections are shown. The green border outlines the result of a simple probability map thresholding at value $p = 0.5$, the yellow border outlines the result of the proposed probability map post-processing, and for the reference, the magenta border outlines the manual segmentation. The bottom row shows three-dimensional models of the same PG: the manual segmentation (left), the result of probability-map thresholding (middle) and the result of post-processing (right). All models are viewed from the same perspective.

Figure 3. Box plots of the segmentation results for the left (first 3 columns) and the right (last 3 columns) parotid gland (PG). Columns *Elastix* and *ANTs* show the results for the proposed multimodal segmentation method, where image pairs were registered using the Elastix and ANTs toolbox, respectively. Columns *CT only* show the results for the monomodal segmentation method, which uses CT images only. Measurements from above to bottom are the Dice coefficient (Dice) and average symmetric surface distance (ASSD).

Figure 4.

A case with visually good registration in the region of parotid glands (PGs), which yielded improved segmentation results (Dice coefficient +5.7% for the left PG and +9.1% for the right PG, using Elastix, and +4.3% for the left PG and +8.5% for the right PG, using ANTs). The magenta (solid) line represents the border of manual delineation of PGs in the reference domain and is therefore located at the same position in all three images. Yellow (dashed) lines represent the results of automatic segmentation methods: on CT slices, results of monomodal segmentation are shown, and on MR slices the results of multimodal segmentation method are shown, where the corresponding registered MR image was used.

Figure 5.

A case where registration was not perfect in the region of parotid glands (PGs), but multimodal segmentation method managed to improve results nevertheless (Dice coefficient +5.0% for the left PG and +8.1% for the right PG, using Elastix, and +4.6% for the left PG and +7.4% for the right PG, using ANTs). The magenta (solid) line represents the border of manual delineation of PGs in the reference domain and is therefore located at the same position in all three images. Yellow (dashed) lines represent the results of automatic segmentation methods: on CT slices, results of monomodal segmentation are shown, and on MR slices the results of multimodal segmentation method are shown, where the corresponding registered MR image was used.

Figure 6.

A case with bad registration and non-improved segmentation results (Dice coefficient -0.4% for the left parotid gland (PG) and -1.6% for the right PG, using Elastix, and -0.4% for the left PG and -1.3% for the right PG, using ANTs). The magenta (solid) line represents the border of manual delineation of PGs in the reference domain and is therefore located at the same position in all three images. Yellow (dashed) lines represent the results of automatic segmentation methods: on CT slices, results of monomodal segmentation are shown, and on MR slices the results of multimodal segmentation method are shown, where the corresponding registered MR image was used.

Table 1

Choice of parameters for registration with Elastix.

Parameter	Rigid registration	Non-rigid registration
Mutual information		
Number of samples used in each iteration	6000	4000
Bins per histogram dimension	32	32
Multiresolution strategy		
Standard deviations $\sigma_x \times \sigma_y \times \sigma_z$ of Gaussian smoothing ($s_x \times s_y \times s_z$ is image voxel size)		
1, level	$4s_x \times 4s_y \times 2s_z$	—
2, level	$2s_x \times 2s_y \times 1s_z$	$2s_x \times 2s_y \times 1s_z$
3, level	$1s_x \times 1s_y \times 1s_z$	$1s_x \times 1s_y \times 1s_z$
Optimisation		
Max, number of iterations	2000	1000

Table 2

Choice of parameters for registration with ANTs.

Parameter	Rigid registration	Non-rigid registration
Mutual information		
Ratio of samples used in each iteration	25%	25%
Bins per histogram dimension	32	32
Multiresolution strategy		
Standard deviation σ of Gaussian smoothing	3/2/1/0 mm	3/2/1/0 mm
Downsampling factors	4/2/2/1	6/4/2/1
Optimisation		
Max, number of iterations	1000/500/250/100	100/70/50/20
Tolerance (last 10 iterations)	10^{-6}	10^{-6}
Gradient step	0.1	0.1

Table 3

Comparison of multimodal parotid gland (PG) segmentation methods using different registration toolboxes (Elastix and ANTs) for image pair registration, and monomodal PG segmentation method (CT only). Row *A* vs. *B* shows in how many cases the segmentation method *A* outperformed the segmentation method *B* in terms of the Dice coefficient and average symmetric surface distance (ASSD).

	<u>Left PG 43 cases</u>		<u>Right PC 42 cases</u>	
	Dice	ASSD	Dice	ASSD
Elastix vs. CT only	33	19	36	22
ANTs vs. CT only	33	17	35	22
ANTs vs. Elastix	21	18	23	22