HHS Public Access

Author manuscript

Med Phys. Author manuscript; available in PMC 2021 December 17.

Published in final edited form as:

Med Phys. 2021 December ; 48(12): 7806–7825. doi:10.1002/mp.15308.

SOMA: Subject-, object-, and modality-adapted precision atlas approach for automatic anatomy recognition and delineation in medical images

Jieyu Li[1,2], **Jayaram K. Udupa**[2], **Dewey Odhner**[2], **Yubing Tong**[2], **Drew A. Torigian**[2]

[1]Institute of Image Processing and Pattern Recognition, Department of Automation, Shanghai Jiao Tong University, Shanghai, China

[2]Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract

Purpose: In the multi-atlas segmentation (MAS) method, a large enough atlas set, which can cover the complete spectrum of the whole population pattern of the target object will benefit the segmentation quality. However, the difficulty in obtaining and generating such a large set of atlases and the computational burden required in the segmentation procedure make this approach impractical. In this paper, we propose a method called SOMA to select subject-, object-, and modality-adapted precision atlases for automatic anatomy recognition in medical images with pathology, following the idea that different regions of the target object in a novel image can be recognized by different atlases with regionally best similarity, so that effective atlases have no need to be globally similar to the target subject and also have no need to be overall similar to the target object.

Methods: The SOMA method consists of three main components: atlas building, object recognition, and object delineation. Considering the computational complexity, we utilize an all-to-template strategy to align all images to the same image space belonging to the root image determined by the minimum spanning tree (MST) strategy among a subset of radiologically near-normal images. The object recognition process is composed of two stages: rough recognition and refined recognition. In rough recognition, subimage matching is conducted between the test image and each image of the whole atlas set, and only the atlas corresponding to the best-matched subimage contributes to the recognition map regionally. The frequency of best match for each atlas is recorded by a counter, and the atlases with the highest frequencies are selected as the precision atlases. In refined recognition, only the precision atlases are examined, and the subimage matching is conducted in a nonlocal manner of searching to further increase the accuracy of boundary matching. Delineation is based on a U-net-based deep learning network, where the original gray scale image together with the fuzzy map from refined recognition compose a two-channel input to the network, and the output is a segmentation map of the target object.

Correspondence: Jayaram K. Udupa, Medical Image Processing Group, Department of Radiology, 3710, Hamilton Walk, 6th Floor, Rm 602W, Philadelphia, PA 19104, USA. jay@pennmedicine.upenn.edu.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

Results: Experiments are conducted on computed tomography (CT) images with different qualities in two body regions – head and neck (H&N) and thorax, from 298 subjects with nine objects and 241 subjects with six objects, respectively. Most objects achieve a localization error within two voxels after refined recognition, with marked improvement in localization accuracy from rough to refined recognition of 0. 6–3 mm in H&N and 0. 8–4. 9 mm in thorax, and also in delineation accuracy (Dice coefficient) from refined recognition to delineation of 0. 01–0. 11 in H&N and 0. 01–0. 18 in thorax.

Conclusions: The SOMA method shows high accuracy and robustness in anatomy recognition and delineation. The improvements from rough to refined recognition and further to delineation, as well as immunity of recognition accuracy to varying image and object qualities, demonstrate the core principles of SOMA where segmentation accuracy increases with precision atlases and gradually refined object matching.

Keywords

anatomy recognition; multi-atlas segmentation; precision atlas selection

1 | INTRODUCTION

1.1 | Background

Prior knowledge-based anatomy segmentation methods have shown their strength and robustness in the field of medical image analysis. Typical methods include those based on shape and geographic models,[1–5] atlases,[6–11] and deep neural network models.[12–16] The models, whether generated directly from the consensus of shapes among a set of samples or based on deep learning networks, all suffer from the problem that information is blurred when models are created, affecting decision making. Compared to shape and geographic model-based methods, which first determine models and then match target objects, some atlas-based methods directly use raw intensity images for decision making on specific patient images, which is the basis of *precision medicine*.

Most applications of the multi-atlas segmentation (MAS) approach[10] lie in automatic structural segmentation in brain magnetic resonance imaging (MRI) data,[17–23] while MAS has also shown usefulness in the segmentation of objects in images of different modalities[24] such as MRI, computed tomography (CT), ultrasonography, and in different body regions such as head and neck (H&N),[25–30] thorax,[31] abdomen,[32,33] and multiple body regions.[34,35] An implicit assumption in the MAS method is that there should be a large enough atlas set, which has complete and perfect segmentations of target objects and which covers the object shape and geographic layout patterns and image intensity appearance patterns of the whole population of subjects under study. Also, when segmenting an object from an unseen input image, it assumes that there exists a subset of the atlases that closely resembles the pattern for this specific input case.[36] These assumptions are usually not satisfied, which result in suboptimal segmentation. It is also unrealistic to obtain an infinitely large atlas set with complete reference delineations of objects that can account for all kinds of individual variations.[37] The basic question of the minimum number of atlas images needed to be able to cover the subject-specific patterns of variation is only now beginning to be addressed.[38]

To further develop the precision idea as applicable to MAS methods, we propose an atlas selection approach named SOMA, utilizing subject-, object-, and modality-adapted precision atlases, which largely increases the implicit patterns included in the atlas set by recognizing different parts of the target object in a novel image from different atlases, where the atlases with the highest frequencies of partial similarity comprise the sample-specific precision atlases.

1.2 | Related works

The basic steps of atlas-based segmentation methods include registration, atlas selection, and label fusion, and numerous strategies are proposed to improve one or more of them specifically adapted to their application and increase the segmentation accuracy. Registration is a fundamental preprocessing step through which the target image and atlas images are adjusted into a same image space where the atlas labels can be spatially propagated properly. Registration can be group-wise[39] via tree-based strategies[40,41] or template strategies,[42] and can be target-specific[43] where all atlases are registered to the target image. Based on the computational complexity, registration can be a simple rigid transformation or a nonrigid deformation.[10,44] Besides simply taken as a preprocessing step, registration can further provide evidence where deformation can be taken as a similarity measure of ranking atlases for atlas selection and assigning weights for label fusion.[45,46]

Proper strategies of atlas selection will help to improve the computational efficiency and the segmentation quality.[44] In Ref[36] the best segmentation quality was estimated by the extreme value theory under the assumption of a given large enough atlas set (up to 5000 atlases). Although larger atlas sets can contain more patterns for tolerating individual variations, computational burden caused by registration[47] could not be afforded in clinical practice. Moreover, the quality of the atlases fundamentally influences segmentation quality,[48] and therefore atlas selection[46,49] is introduced into the atlas-based segmentation methods to yield more accurate segmentation and to reduce the computational load of registration.[50]

Atlas selection can be and is usually conducted in an offline manner before considering the target image.[44] In Ref[50] a strategy was provided of initially clustering all atlases and then choosing the most representative cluster to fully register to the target image. The selection can also be conducted for each specific target image in an online manner[44] according to image-based similarity metrics such as intensity-based metrics, features, and degree of overlap,[51] and/or meta-information such as patient age and gender.[17] In Ref[52] a generic algorithm was proposed, where the intensity-based metric and Dice coefficient (DC) are used to measure similarity in a two-stage atlas selection process. In Ref[17] rigid and all-to-template registration is conducted to align all images into a same image space, and after atlas selection, nonrigid and target-specific registration is conducted on the most similar atlases to reach a higher accuracy of alignment. A hierarchical strategy is used in Ref[32] where registration, atlas selection, and atlas weighting are sequentially refined at the global, organ, and voxel levels. As opposed to selecting target-specific atlases, Ref[21] selected representative atlases in the low-dimensional data space via a sparsity-based strategy.

Label fusion is another key element of the MAS approach. The most straightforward strategy for label fusion is majority voting,[53] where labels on target voxels are determined

by the most common agreement from atlases. As an evolution of majority voting, intensity-based,[32] deformation-based, and overlappingbased similarity measures are also often used to determine weights in voting. Another series of commonly used strategies are expectation-maximization-based STAPLE[54] and its extensions,[48,55] which introduce probabilistic models into label fusion.

The explosive use of machine learning and deep learning strategies also contribute to atlas-based methods in different applications, including image registration and label propagation,[56] atlas ranking and selection,[33,51,57] and feature extraction and/or label fusion.[58–61]

Although numerous strategies have been proposed to improve performance of atlas-based segmentation, and the importance of atlas selection has been emphasized from different aspects, the criteria for atlas ranking are all purely based on different kinds of similarity measures proposed in the literature. Besides, current methods are mainly based on the statistical decision from all available or selected atlases, where the pattern information from each single atlas is mixed up and is not fully utilized. In this work, we propose the SOMA approach starting from a novel viewpoint that different parts of a target object in the novel image can be recognized (matched) by different atlases, and that the frequencies of regional best match, instead of similarity itself, will be an effective strategy in selecting precision atlases. Furthermore, this strategy can be employed recursively to refine the “precision” of the atlases.

An early version[62] of this work has been published in the SPIE 2021 Medical Imaging conference. In the present paper, we make several major extensions:(i) The conference version focused on the recognition of organs in H&N region, whereas the present work contains, in addition to the H&N organs, the organs in the thorax region that have more varying shapes, sizes, and intensity and contrast distributions. (ii) Deep learning-based delineation (DLD) is conducted based on recognition maps, and thus the complete improvement of boundary interpretation can be observed through the whole process of anatomy segmentation. (iii) Extended experiments are conducted to demonstrate the effectiveness of the selected recognition parameters. (iv) A full presentation of methods, results, discussions, and background literature is also contained in this work.

1.3 | Outline of approach

The SOMA approach is depicted in Figure 1 and is described in detail in Section 2. There are three main components in this method: atlas building, object recognition, and object delineation. In atlas building, we align all atlas images into a unified image space, which belongs to a template image determined by the minimum spanning tree (MST) algorithm[40] among a set of preselected radiologically near-normal images. Then, a two-stage recognition process, involving rough recognition (RoR) and refined recognition (ReR), is conducted to generate fuzzy maps for object localization. Only the atlas images with subimage-level best match contribute to the membership map for recognition in the same subimage region. In RoR, all images in the atlas set are examined on each pixel-centered subimage, and the frequency of best match is counted for each of the atlases. The atlases with highest frequency are selected as precision atlases and utilized in ReR. Refined region of interest

(ROI) and nonlocal searching are also applied to generate fuzzy membership maps with better localization and more precise boundary matching. Lastly, the fuzzy map from ReR is further refined to the delineation mask via a deep learning model, where the original gray scale image and the fuzzy map compose a two-channel input to a U-net[63] based network, and the output is the delineation mask for the target object.

Section 3 describes experiments conducted for verifying the SOMA method by the datasets of CT images in the H&N and thorax body regions from the Hospital of the University of Pennsylvania. Comparisons, gaps remaining in this work, and avenues for potential improvements are discussed in Section 4. Our conclusions are given in Section 5.

2 | METHOD

Notation:

B: Human body region studied.

m: Number of image modalities considered.

$N_1, N_2, \ldots, N_m$: Number of images in modalities 1, …, m, respectively, available for atlas building.

$\mathcal{O} = \{O_1, \ldots, O_L\}$: L objects considered in body region B.

$\mathcal{I}^a = \left\{I_1^a, \ldots, I_{N_1}^a, I_{N_1+1}^a, \ldots, I_{N_1+N_2}^a, \ldots, I_{N_1+\cdots+N_m}^a\right\}$: A set of images of body region B available from m modalities.

$\mathcal{I} = \left\{I_1, \ldots, I_{N_1}, I_{N_1+1}, \ldots, I_{N_1+N_2}, \ldots, I_{N_1+\cdots+N_m}\right\}$: Images of $\mathcal{I}^a$ after they have been registered to a template determined by the MST algorithm.[40]

$\mathcal{J}^\ell = \left\{J_1^\ell, \ldots, J_{N_1}^\ell, J_{N_1+1}^\ell, \ldots, J_{N_1+N_2}^\ell, \ldots, J_{N_1+\cdots+N_m}^\ell\right\}$: Binary images representing true segmentations of object O_ℓ in the images in $\mathcal{I}$. Note that when segmentations are obtained from images in $\mathcal{I}^a$, the same registration operations applied to images in $\mathcal{I}^a$ to produce $\mathcal{I}$ are assumed to have been applied to these segmented binary images to obtain $\mathcal{J}^\ell$.

For simplicity, below we will assume that the number of modalities $m = 1$ and that $N_1 = N$. All that is described generalizes readily to the case of $m > 1$. With these assumptions, let $\mathcal{I}$ and $\mathcal{J}^\ell$ be defined as $\mathcal{I} = \{I_1, \ldots, I_N\}$, $\mathcal{J}^\ell = \left\{J_1^\ell, \ldots, J_N^\ell\right\}$, $\ell = 1, \ldots, L$.

$V_{\omega,I}(v)$: A $\omega \times \omega$ 2D subimage centered at pixel v of I.

The SOMA approach consists of an initial atlas building step, which is followed by object recognition and delineation steps.

2.1 | Atlas building

The atlas set is built by aligning all atlas images into the same image space, and the corresponding binary masks of target objects are geometrically transformed in the same

manner. For the target image under investigation, it should also be transformed to the same image space before the subsequent processes of recognition and delineation are performed. Considering the computational limitations and the time-consuming nature of the problem, all-to-template registration is utilized in SOMA, although this does not guarantee global optimality. Specifically, following SOMA's spirit of selecting precision atlases, the intersubject variations are preserved. As such, a seven-parameter transformation is applied in the SOMA approach, where only global shift, rotation, and isotropic scaling are applied to adjust the overall position, pose, and scale of each subject during registration, instead of the nonrigid registration[64] that is commonly used in atlas-based segmentation but which seems to be less effective for anatomical objects in body regions outside of the brain.[10]

To choose a constant template image from the atlas set $\mathcal{I}^a$, we first determine a subset of candidate images of $\mathcal{I}^a$, denoted by $\mathcal{I}^{\mathrm{a}}_R$, which are *radiologically near-normal*, with the least amount of artifacts and pathological abnormalities. The template is determined from $\mathcal{I}^{\mathrm{a}}_R$ by an MST algorithm.[40] A complete weighted directed graph is first established where the nodes are the candidate images in $\mathcal{I}^{\mathrm{a}}_R$ and the arc weights/costs are assigned by the dissimilarity between the node images. Mean absolute difference (MAD) is used as a metric to measure the dissimilarity between two candidate images, or weight for arc ($\mathcal{I}^a_S$, $\mathcal{I}^a_T$), as shown in Equation (1):

$$w\left(I^a_S, I^a_T\right) = \frac{\sum_{v \in I^r_{Sb} \cup I^a_{Tb}} \left|I^r_S(v) - I^a_T(v)\right|}{\left|I^r_{Sb} \cup I^a_{Tb}\right|}, \tag{1}$$

where the source image I^a_S is registered to the target image I^{a}_T and transformed into I^r_S, and where I^r_{Sb} and I^a_{Tb} represent the binary foreground regions inside of the outer skin boundaries of I^r_S and I^a_T, respectively, to exclude the influence of background information, such as the scanner table, on the dissimilarity measure. After the graph is set up, an MST of the graph with least total cost is found.

The root image I^a_{Root}.of the MST is used as the template target image for the registration of all other images, including all atlases and future coming test images. In this way, all images are registered to the same image space with unified global position, pose, and scale. The set of atlases with obvious pathological abnormalities or artifacts is denoted by $\mathcal{I}^a_A$ Zero-padding in the z (craniocaudal) direction is necessary so that after registration, the images are properly and consistently represented in all studies. Otherwise, studies with shorter superior to inferior dimension may be cut off after registration at their ends in the craniocaudal direction. After registering all images to the root image, we will have sets $\mathcal{I}$, $\mathcal{I}_R$, and $\mathcal{I}_A (\mathcal{I} = \mathcal{I}_R + \mathcal{I}_A)$ corresponding to sets $\mathcal{I}^{\mathrm{a}}$, $\mathcal{I}^{\mathrm{a}}_R$, and $\mathcal{I}^{\mathrm{a}}_A (\mathcal{I}^{\mathrm{a}} = \mathcal{I}^{\mathrm{a}}_R + \mathcal{I}^{\mathrm{a}}_{\mathrm{A}})$, respectively. Although the root image I^a_{Root} does not change after the entire registration process, to make notation uniform, we will denote it simply by I_{Root}. The subset $\mathcal{I}_R$ is used for estimating the parameters of the SOMA-R approach, while the whole set $\mathcal{I}$ is employed for building the atlas.

2.2 | Object recognition

In the SOMA approach, objects are recognized one at a time. The SOMA recognition procedure, SOMA-R, is composed of two stages, RoR and ReR. In the RoR stage, an object O_ℓ is localized (recognized) in a given image I by examining all atlas images in $\mathcal{I}$ and identifying an atlas subset P from $\mathcal{I}$ that can be best associated with O_ℓ in I. In the ReR stage, the locality of O_ℓ is sharpened by examining the atlas images in only P.

As mentioned previously, different parts of the segmentation can come from different atlases. Only the atlas with best local similarity with the target image contributes to the recognition map of the target object O_ℓ and is determined as the precision atlas in this local region. The similarity is locally measured in sliding $\omega \times \omega$ 2D windows. The frequency of local best match over the image domain is the measure used to determine the overall precision atlases for O_ℓ

For describing SOMA-R, we will slightly modify the representation of binary images in $\mathcal{I}^\ell$ by changing background voxel values 0 to −1, but will still maintain the binary representation. The reason for making this change is that we wish to add up the contributions from the object parts (represented by voxels with value 1) and background parts (represented by voxels with value −1) from all precision atlases for each voxel v of I to develop a fuzzy map of O_ℓ Correspondingly, the subimage $V_{\omega,I}(v)$ will also be comprised of only elements 1 and −1.

2.2.1 | Procedure SOMA-R

Input: A test image I^a, atlas set $\mathcal{I}$ and binary images $\mathcal{I}^\ell$, $\ell = 1,\ldots, L$ after registration; an image similarity function ψ (sum of squared difference [SSD] in this procedure); a threshold θ for the similarity function ψ; a subimage size ω; a ratio δ% of the precision atlases selected for ReR with respect to the whole atlas set; and a nonlocal floating window searching range f_r.

Auxiliary variables: Atlas maps AM_ℓ for RoR and am_ℓ for ReR for recording the atlas index of the local best-match atlas; counters $C(n)$ and $c(n)$ for counting the frequency of atlas image $I_n \in \mathcal{I}$ selected as the local best-match atlas in RoR and ReR, respectively.

Output: Fuzzy membership maps FM_ℓ and fm_ℓ corresponding to RoR and ReR, respectively.

Begin:

R0. Register I^a to the template root image I_{Root}. Let the transformed version of I^a be I.

For each object O_ℓ *do*

R1. Determine an initial ROI, denoted R_{in}, by dilating the union of the foreground regions of the images in $\mathcal{I}^\ell$.

Rough recognition:

R2. Set all voxels of FM_ℓ and AM_ℓ to 0 and so also all elements of $C_\ell(n)$, $n = 1, \ldots, N$.

R3. *For* each voxel v of I inside region R_{in} *do*

R4. Determine subimage $V_{\omega,I}(v)$ at v.

R5. Determine image $I^* \in \mathcal{I}$ such that $I^* \in \arg\min_{K \in \mathcal{I}} \{\psi(V_{\omega,I}(V), V_{\omega,K}(V))\}$. Let $J^* \in \mathcal{I}^\ell$ be the binary image representing O_ℓ in I^*.

R6. *If* $\psi(V_{\omega,I}(v), V_{\omega,I^*}(v)) \le \theta$, *then* add $V_{\omega,I^*}(v)$ to FM_ℓ and set value of $AM_\ell(v)$ to the index associated with I^*.

R7. *EndFor*

R8. Threshold the fuzzy map FM_ℓ into a binary map BM_ℓ

R9. Determine a refined ROI, R_{re}, by dilating foreground region of BM_ℓ

R10. *For* each voxel v of I inside region R_{re} *do*

R11. Increment counter $C_\ell(AM_\ell(v))$ by 1.

R12. *EndFor*

R13. Normalize and output FM_ℓ and output $C_\ell(n)$, $n = 1, \ldots, N$.

Refined recognition:

R14. Set all voxels of fm_ℓ and am_ℓ to 0 and so also all elements of $c_\ell(n)$, $n = 1, \ldots, N$.

R15. Rank $C_\ell(n)$, $n = 1, \ldots, N$, in descending order. Top $\delta\% \times N$ atlases with highest counter values $C_\ell(n)$ compose the precision atlas set $\mathcal{I}_P$ and $\mathcal{I}_P^\ell$ for I and object O_ℓ

R16. *For* each voxel v of I inside region R_{re} *do*

R17. Determine subimage $V_{\omega,I}(v)$ at v.

R18. Determine nonlocal floating window searching range $R_f(v)$ at v.

R19. Determine image $I^* \in \mathcal{I}_P$ and nonlocal best-match position $v^* \in R_f(v)$ such that $I^*, v^* \in \arg\min_{K \in \mathcal{I}, V' \in R_f(v)} \{\psi(V_{\omega,I}(v), V_{\omega,K}(v'))\}$. Let $J^* \in \mathcal{I}_P^\ell$ be the binary image representing O_ℓ in I^*.

R20. *If* $\psi(V_{\omega,I}(v), V_{\omega,I^*}(v^*)) \le \theta$, *then* add $V_{\omega,I^*}(v^*)$ to fm_ℓ on the subimage with the center v and set value of $am_\ell(v)$ to the index associated with I^*.

R21. *EndFor*

R22. *For* each voxel v of I inside region R_{re} *do*

R23. Increment counter $c_\ell(am_\ell(v))$ by 1.

R24. *EndFor*

R25. Normalize and output fm_k and output $c(n)$, $n = 1, \ldots, N$.

EndFor

End

Details of the SOMA-R procedure are as follows: In *Input*, the SSD is used as the function ψ to evaluate similarity (dissimilarity) between subimages of I and atlas images in $\mathcal{I}$. Pearson correlation coefficient (PCC) and normalized mutual information[17,52] (NMI) may also be suitable similarity metrics in different situations. As we use a dissimilarity function, the local best-match atlas should have the least SSD, and the threshold θ is used to avoid cases where no atlas is locally similar to the target subimage.

Atlas maps AM_ℓ and am_ℓ and counters $C(n)$ and $c(n)$ are auxiliary variables for determining the precision atlas set P. After RoR, the initial ROI (R_{in}) is refined into R_{re} (target specific), and the counters are generated from atlas maps after excluding counts in the unrelated region. While $C(n)$ can produce nonzero counts for any atlas image $I_n \in \mathcal{I}$, $c(n)$ will produce nonzero counts only for $I_n \in \mathcal{I}_P$. When the whole recognition process is iteratively refined, the ROI has the potential to undergo continued refinement based on the counters and fuzzy maps.

The outputs of SOMA-R are fuzzy maps FM_ℓ and fm_k which map the location of object O_ℓ roughly as a fuzzy mask over image I. The map values $FM(v)$ and $fm(v)$ at voxel v indicate the cumulative votes of membership on v from all best-match atlas subimages going through v. This is not intended to be a precise delineation of O_ℓ but instead a rough indicator of the whereabouts of O_ℓ in I.

Steps R3–R7 compose the core of RoR where, at each voxel v, first a $\omega \times \omega$ 2D subimage $V_{\omega,I}(v)$ of I centered at v is found (R4), and then a homologous subimage $V_{\omega,I^*}(v)$ over all atlas images in I that best matches with $V_{\omega,I}(v)$ as per the similarity function ψ is determined (R5). If this match is at or above a certain confidence level ($\psi(.) \leq \theta$), then the evidence for the location of O_ℓ at v in FM_ℓ is updated, and $AM(v)$ is also updated with the index of the atlas corresponding to I^* (R6). The updating of FM_ℓ is accomplished by adding to the current FM_ℓ map the entire subimage $V_{\omega,J^*}(v)$ of the binary image J^*, corresponding to I^*, with all of its −1 and 1 values (see Figure 2). At the end of this loop (R7), two outcomes are expected: a rough location of O_ℓ in I to emerge in the FM_ℓ map, and an atlas index map AM_ℓ showing the index of the best-match atlas for each voxel location, where $C(n)$ will be accumulated from AM_ℓ for each atlas in R10–R12.

In steps R8–R9, the refined ROI R_{re} is dilated from BM_ℓ to focus on the specific target image I. The fuzzy map FM_ℓ is converted to the binary map BM_ℓ by the threshold value 1, where the FM_ℓ map values can range from $-\omega \times \omega$ to $+\omega \times \omega$, and the threshold value 1 indicates that there is at least one more subimage from best-match atlas that votes on the foreground than on the background. In Step R13, the FM_ℓ map can be normalized to the range [−1, 1] without affecting the 0 values in the map. A desirable property of SOMA-R is that the membership for not just the object, but also for surrounding background, is determined.

ReR starts from Step R14. Although the implementation details are similar to those of RoR, ReR shows its strength in pursuing better localization by considering the refined set of precision atlases, refined target-specific ROI, and the nonlocal best-match searching strategy.

Step R15 constitutes the heart of the ReR strategy. If the atlas building stage has collected enough images to capture within the precision atlas set (and not necessarily in the whole atlas set without the precision atlas concept), the particular object layout and intensity distribution pattern presented in image I, then we expect counters $C(\ell,n)$ to yield evidence of images that maximally match with I for O_ℓ. The counters show the frequency of each atlas being the best match atlas for subimages, which imply that atlases are not necessarily overall similar to I for O_ℓ but are frequently similar in parts. The atlases are ranked according to the counters, and the top δ% atlases among the whole atlas set compose the precision atlas set $\mathcal{I}_P$, in which the atlas quality is outstanding compared to remaining atlases in $\mathcal{I}$, and the corresponding binary masks are denoted by $\mathcal{J}_P^\ell$.

In Step R19, a nonlocal searching strategy is used to alleviate regional individual differences and misregistrations, which are difficult to consider in imagelevel registration. The target subimage $V_{\omega,I}(v)$, like a floating window, searches the best-match atlas simultaneously within the 3D searching range $R_I(v)$ and among the atlases in $\mathcal{I}_P$. Let f_r refer to the radius of the searching range in voxels in the dimension with lowest resolution where r refers to the ratio between slice spacing and 2D pixel length. The searching range should be an isotropic region with $(2r \times f_r + 1) \times (2r \times f_r + 1) \times (2 \times f_r + 1)$ voxels centered on the target position v. Given a typical situation where the voxel size in a CT image is $1 \times 1 \times 2$ mm and f_r is set as 2, the searching of best match should be restricted inside a range of $9 \times 9 \times 5$ voxels.

2.2.2 | Parameter determination—Four parameters are involved in the SOMA-R process, namely the threshold θ for the (dis)similarity function ψ, the subimage size ω, the ratio δ% for selecting precision atlases, and the nonlocal floating window searching range f_r. Among these parameters, the threshold θ and the window size ω are object-dependent parameters, and the ratio δ% and the searching range f_r are empirically decided upon according to the representability of the atlases. Intuitively, when the atlas set has perfect representability, the target object in the specific test image can be well represented by very few atlases. Conversely, the ratio δ% should be large if the atlas set does not contain that many patterns, such that the best-match subimages scatter widely among different atlases. If only a limited atlas set is available, the searching range f_r should also be large to provide more chances for subimage matching, although a large f_r may also lead to the problem of mismatching with surrounding confounding objects.

The threshold θ and the window size ω are object-specifically decided by experiments using the near-normal atlas subset $\mathcal{I}_R$. A leave-one-out strategy is used in RoR with different combinations of θ and ω. The combination yielding the best average DC on binary masks BM_i is utilized in the actual SOMA-R procedure.

2.3 | Object delineation

If we binarize fm_ℓ into bm_ℓ we can observe that bm_ℓ approaches a delineation mask after utilizing precision atlases and the floating window strategy. However, it will still show scattered points, which are located appropriately in the vicinity of the location of object O_ℓ in I, but which cannot contour the object accurately. Thus, postprocessing is needed. Deep learning-based methods are under explosive development in semantic segmentation in medical images, and U-net is one of the most used fully convolutional end-to-end networks. In the SOMA delineation (SOMA-D) procedure, we use a 2D U-net based network, where inputs are the two-channel images composed of the original gray scale image I and the fuzzy map fm_ℓ output by SOMA-R ReR, and output is the corresponding semantic segmentation mask for O_ℓ

2.3.1 | Network architecture—The network architecture is illustrated in Figure 3. Binary cross-entropy is taken as the loss function and Adam optimizer is used. Batch normalization is conducted, and batch size is determined according to ROI size and memory capacity. ReLUs in the encoder path are leaky with slope 0.2, and ReLUs in the decoder path are not leaky.[65]

As in the definition of R_{in}, the network input is trimmed by an ROI determined from fm_ℓ of training samples. The bounding boxes of voxels where $fm_\ell(v) > -0.4$ for each training sample is first determined. Then, a larger box fully covering all of the bounding boxes is taken as a proper 2D ROI. To satisfy the input size of the network, where there are three convolutional layers with stride 2, the ROI is further expanded to a slightly larger size of multiples of eight. As all images are aligned to the same image space at the beginning of the SOMA method, the target object will be contained properly inside this ROI with very high likelihood. This also improves delineation specificity of the U-net.

2.3.2 | Training images—Atlas images in $\mathcal{I}$ are employed for training, where a leave-one-out strategy is used in the complete SOMA-R procedure to generate the ReR map fm_ℓ for them. Each atlas image is taken as the target image while other images are taken as atlases. Slices of the fuzzy map fm_ℓ together with the original intensity slices in I are trimmed by ROI and concatenated into two-channel 2D inputs to the network. Data augmentation is conducted to mimic different recognition qualities by shifting and strengthening (sharpening) or weakening (blurring) the fuzzy maps for localization error (LE) and scaling error (SE) as in Equations (2) and (3), respectively.

$$fm_\ell^{LE}(v) = fm_\ell(v'), \tag{2}$$

$$fm_\ell^{SE}(v) = \max(\min(fm_\ell(v) \pm p, 1), -1), \tag{3}$$

where in Equation (2), v' is spatially shifted from the original voxel v with deviation s, that is, $v = (i, j, k)$, $v' = (i \pm r \times s, j \pm r \times s, k)$ or $v' = (i, j, k \pm s)$, and r shows the ratio between slice spacing and pixel width as in determining the floating window range in the section describing ReR. The deviation s is also able to mimic the potential error, which cannot be overcome by or is introduced by floating windows. In Equation (3), p stands for

the membership value added to or subtracted from the original fuzzy membership value $fm(v)$, and intuitively indicates the situation where more or less atlases among neighboring $\omega \times \omega$ atlases agree with the membership of voxel v as foreground.

2.3.3 | Testing images—For testing images, fuzzy maps and the original images are trimmed into an ROI using the same size and position as for the training samples. The output is the segmentation mask with the trimmed ROI size, which can then be restored back to the original image size.

3 | EXPERIMENTAL RESULTS

3.1 | Datasets and experiments

3.1.1 | Datasets—This retrospective study was conducted following approval from the Institutional Review Board at the Hospital of the University of Pennsylvania (HUP) along with a Health Insurance Portability and Accountability Act waiver. Experiments were conducted on CT images of two body regions, H&N and thorax, from 298 and 241 patients, respectively. The routine clinically acquired images are for radiation therapy planning of patients with cancer in the two body regions. Nine objects in the H&N region and six objects in the thorax region as defined in[66] are considered: CtEs, CtSC, Mnd, OHPh, SPGLx, RPG, LPG, RSmG, and LSmG in the H&N region; TSC, TEs, TB, Hrt, RLg, and LLg in the thorax region. The full names and acronyms for these objects are listed in Table 1 for ease of reference. Object-level quality (OQ) is manually evaluated by experts in terms of whether the object and its surrounding tissue are involved by pathology or whether the imaging quality is affected by artifacts,[67] based upon which the object samples are divided into groups of good quality (GQ) and poor quality (PQ). Thirty-six subjects in the H&N region and 39 subjects in the thorax region show overall GQ on all considered objects and hence are selected as radiologically near-normal images comprising set $\mathcal{I}_R^a$. As described in[66], a GQ study of an object contains deviations due to artifacts, abnormalities, and so forth in not more than three slices through the object, and a study that is not GQ is considered PQ for that object.

The voxel size varies from $0.93 \times 0.93 \times 1.5$ to $1.6 \times 1.6 \times 3$ mm^3. The root images determined by SOMA have a resolution of $1 \times 1 \times 3$ and $0.97 \times 0.97 \times 3$ mm^3 in H&N and thorax regions, respectively, and the sizes of all images are unified to $512 \times 512 \times 92$ and $512 \times 512 \times 128$ voxels, respectively, after registration.

3.1.2 | Experiments—The SOMA method is N-fold cross-validated on samples in the whole datasets excluding the near-normal set $\mathcal{I}_R$, from which parameters are determined for SOMA-R as explained above. As not all samples are with complete reference masks of all considered objects, the division of folds is also different as shown in Table 1, where N_F is the number of folds, A_{FR} denotes the number of atlases contained in $\mathcal{I}$ in each fold, and T_{eF} stands for number of test samples in each fold.

There are four parameters contained in the SOMA-R procedure, which include 2D subimage size ω (in pixels), similarity threshold θ, ratio δ% for precision atlas selection, and f_r for

floating window searching, among which $\delta\% = 20\%$ and $f_r = 2$ are empirically determined according to the representability of the atlases, and ω and θ are experimentally determined for each object by testing different combinations of them on $\mathcal{I}_R$. θ was selected from the values of {200, 400, 800, 1200}, and ω was initially selected from the values of {5, 11, 17} for objects in the H&N region and from {23, 33, 43, 53, 63} for objects in the thorax region. These initial candidate ω values are determined based on object thickness in different body regions (objects in H&N are generally much smaller than those in thorax), and larger values are tested until the highest DC has been reached for BM_i in $\mathcal{I}_R$. Parameters obtained for all considered objects are listed in Table 2.

To quantitatively assess the performance of the SOMA method, we analyze the LE and SE for the RoR and ReR results, and DC and average symmetric distance[68] (ASD) for the delineation results. Although FM_i and fm_i are fuzzy maps, they do not represent the probability values of image voxels belonging to the object or background, but instead show the membership from agreement over all atlases, and so the binary masks BM_i and bm_i are used in evaluation. LE is defined as the distance between geometric centers of the reference mask and BM_i or bm_i; SE is the ratio of the recognized object size to its true size. The size of an object represented by the binary mask is calculated by the root of the sum of eigenvalues corresponding to the principal components of the object.[5] LE and ASD are measured in millimeters, and SE and DC are unitless, where cases with perfect overlap should show 0 mm for LE and ASD, and 1 for SE and DC.

3.2 | Results

3.2.1 | Image examples—Image examples are illustrated in Figures 4 and 5 for all considered objects in the two body regions separately. Reference masks (first column), fuzzy maps from RoR and ReR procedures (second and third columns), and delineation masks (fourth column) are overlaid on 2D slices of gray scale images (first row) and overlapped by reference contours (second row). The corresponding surface renditions (for binary masks) and fuzzy volume renditions (for fuzzy masks) are shown as well (third row). From the comparisons of the results from recognition to delineation, we observe gradual improvement, including improvement from RoR to ReR, where the fuzziness of the membership maps is reduced and the interpretation of boundaries is improved, as the latter takes advantage of more precision atlases for the specific target object sample and the better matching introduced by the nonlocal floating window strategy; and from ReR to delineation, where the fuzziness is further reduced and binary masks are produced.

It should be noted how RoR captures the whereabouts of the objects quite sharply and how ReR already appears to demonstrate delineation, albeit fuzzily, quite well. The details captured by ReR are well portrayed in the fuzzy volume renditions, especially for objects with subtle surface details like for CtSC, CtEs, Mnd, SpGLx, TB, TSC, and TEs, sometimes notwithstanding the accompanying false positive regions.

3.2.2 | Quantitative evaluation—Quantitative evaluation results are summarized in Tables 3 and 4, where results of samples with different object quality (OQ) are separately evaluated in terms of RoR, ReR, and DLD.

We make the following observations from the results shown in the tables.

i. Having observed gradual improvement in the results moving from RoR to ReR and to delineation for the image samples shown in Figures 4 and 5, we observe a similar improvement in the quantitative results in terms of decreasing LE values and bringing SE closer to 1 for recognition, and increasing DC values for delineation. By considering only precision atlases and utilizing a floating window searching strategy, the ReR advances from RoR in a manner of better boundary matching. Most of the improvements on results from RoR to ReR and from ReR to delineation are statistically significant with p-value <0.05. Only OHPh, RPG with GQ, RLg with GQ, and TSC with PQ slightly decreased in mean SE from RoR to ReR, and only CtSC with PQ slightly decreased in mean DC from ReR to delineation, although with corresponding decreases in standard deviation.

ii. In RoR, most objects yielded LE around or less than 6 mm, which is twice the unified slice spacing in both body regions (roughly equating two voxels), and the error is further decreased in ReR. The long sparse objects in the thorax region, that is, TB, TSC, and TEs, are more challenging, while TB and TSC are refined toward or under 6 mm in LE. However, TE has a larger LE of up to 14 mm even after ReR. Such a large error may be explained by two reasons: (a) the difficulty in consistently defining the two ends of certain long sparse objects, leading to large errors in the z-direction, and (b) the difficulty in segmenting soft tissue objects with low contrast. Both reasons lead to challenges for segmentation of TEs, especially along its inferior aspect where it joins the stomach at the gastroesophageal junction. Further analysis demonstrates that the average in-plane LE for TEs is 4.489 and 4.523 mm for GQ and PQ samples, respectively, in RoR, which improve to 3.605 and 3.862 mm, respectively, in ReR, showing that the large LE for TEs indeed is mainly attributable to error in the z-direction.

iii. The evaluation results on GQ and PQ samples are similar in all RoR, ReR, and DLD stages, while the model-based method, such as our previously proposed AAR-RT method,[66] shows obvious differences in recognition and delineation results for samples with different qualities as compared in Table 6. This phenomenon indicates that the SOMA method is less influenced by OQ and demonstrates one of the core principles of the precision atlases that the target object sample is only recognized by atlases with local-level best match and will not be influenced by atlases with less similarity. Hence, the samples of various qualities can be well recognized by a sample-specific precision atlas subset if the whole atlas subset can cover different object qualities.

4 | DISCUSSIONS

4.1 | Recognition based on regional similarity

One of the strengths of the proposed method is its ability to recognize different regions of the target object in a novel image via subimage matching with different atlases. Then,

atlases with the highest frequencies of regional best match are selected as the subset of precision atlases. In other words, our method always focuses on each object-level image sample, and that's why we call it a subject-, object-, and modality-adapted method, which is not a progressive strategy sequentially conducted in subject, object, and modality levels, but to consider the three elements as a whole simultaneously. Although we present results only on CT images in this work, the method directly transfers to other modalities or multiple modalities used simultaneously. These extensions will be reported in our future papers.

Figure 6 gives an example of RoR of left lung (LLg) in a novel image to demonstrate that the recognition process should not be conducted at subject level. In the figures, the novel image under consideration is taken as the base (in grey) and overlayed by intermediate results (in orange). Figure 6a–c shows three representative slices of the novel image going in the craniocaudal direction, which are overlayed on the atlas image with least overall SSD. As there is huge anatomic population variation, although the overall relationship and location of anatomy is the same with respect to each person, the subject-level similarity can only guarantee the rough alignment of scale, position, and posture of the whole-body region. Figure 6d shows the initial R_{in}, which is determined by all atlases and taken as the range to conduct subimage matching. Figure 6e is the atlas map (AM) representing the indexes of atlases, which reach regional best match, where the intensities 0–199 represent the 200 atlases under consideration, and the intensity 200 represents the region outside R_{in}. Quantitative statistics show that the 10 atlases with best subject-level similarity rank 29th in average in the frequencies of regional best match, and conversely, the 10 atlases with most frequent best-match rank 28th in average in overall similarity. This demonstrates that subject-level matching is much inferior to regional best match. Figure 6f portrays the detailed recognition process where regions of the target object (LLg) are matched by different atlases based on regional similarities, and the recognition map is generated from binary masks of atlases with regional best match, as shown in Figure 6g.

4.2 | Comparison on different empirical parameters

There are two empirical parameters in the SOMA method, namely, the ratio δ% for selecting precision atlases with the highest frequency of best match and the floating window search range f_r in the nonlocal matching strategy. Another implicit empirical factor is that the ReR is conducted only once in all experiments above. Whether continuous refinement following the current strategy will further improve the recognition accuracy is still under investigation. Experiments on these three empirical parameters are conducted as follows: (i) After RoR and ReR, refinement is continued where ReR_2, ReR_3, and ReR_4 stand for iterative refined recognition with ratio δ_1% = 50%, such that δ_2% = 20% × 50% = 10% for ReR_2, δ_3% = 10% × 50% = 5% for ReR_3, and finally δ_4% = 5% × 50% = 2.5% for ReR_4. At the same time, the ROI is also iteratively refined based on the recognition mask yielded from the previous stage. (ii) Experiments on different δ%, that is, 50% (1/2), 33% (1/3), 10% (1/10), and 5% (1/20) are conducted to check if 20% is a reasonable ratio for refined atlas selection. (iii) Different floating window searching ranges are tested. f_r = 0, 1, 2, and 3 separately denote the radius (in the unit of slice spacing) of the maximum extension from the tested voxel position. As in our experiments the unified resolution of images is around 1 × 1 × 3 mm,

the searching range with, for example, $f_r = 3$, will be a 3D searching range with 19 × 19 × 7 voxels centered on the test voxel.

Experiments are conducted on two typical objects with medium recognition difficulty: a small blob-like object RPG (right parotid gland) and a long sparse object CtEs (cervical esophagus), for which only one fold of test samples, that is, 58 test samples with 150 atlases for RPG and 82 test samples with 200 atlases for CtEs, are contained in the experiments of the empirical parameters. Quantitative evaluation results are shown in Table 5. For each parameter, cases with best DC or LE are marked in bold. Although the selected parameters do not always yield best results, they show no significant difference from the best cases (p-value >0.05). In addition, they are of less computational burden compared to the best cases, where larger δ% and f_r introduce extra computational burden in ReR.

4.3 | Comparison with methods in literature

The same datasets are used in our previous work of the model-based AAR-RT method.[66] AAR-RT was designed to recognize all important organs in the target body region, which is based on high-level anatomic priors including the hierarchy of all organs under consideration and their fuzzy shape models. The hierarchy is a tree that defines the optimal order for recognizing organs and the relative positions and scales of the organ on each offspring node with respect to its parent node. These entities are estimated from a set of near-normal images. Typically, the skin of the body region is taken as the root organ and is first recognized by proper thresholds. Then, other organs are sequentially recognized based on the hierarchy, and refined based on local image intensities. Recognition and delineation are both improved by the proposed SOMA method as shown in Table 6, which quantitatively compares the influence of OQ in recognition and delineation quality via the two methods. OQ is less influential on the SOMA method than on the earlier model-based method, as the target object can be well recognized if its quality is covered in the spectrum of atlas images via SOMA, while model-based methods typically generate object models only based on normal subjects, and recognition may be less accurate for object samples with pathological abnormalities.

We also compared the proposed method with another two typical deep-learning methods. One is utilizing an end-to-end U-net architecture as in Figure 3 without applying SOMA recognition process, that is, the network input is with the original image size (512 × 512 pixels) without ROI cropping via the fuzzy recognition map, and the output is binary mask with the same size as input. The results are shown in the *Baseline* column of Table 6. Comparing with the results of SOMA, although in many cases the DC values show differences within 1%, results for LSmG, Hrt, and TB are largely degraded without the guidance of localization and membership confidence from SOMA recognition.

The other method under comparison utilizes a neural network-based similarity measure to determine the weight of atlases for label fusion. The results are shown in the *Sim_Net* column of Table 6. The similarity network was proposed in Ref[69] and originally targeted for myocardial segmentation in CT and MR images, where the regions around LV myocardium were precropped. The similarity network is designed to map image patches into an embedding space, and the similarity is calculated based on a softmax function over the

Euclidean distance between atlases and target patches. Each training sample is selected around the boundaries of the target object, containing a patch of target image and two atlas patches, which include one positive patch with DC > 0.9 (ground truth similarity 1) and one negative patch with DC < 0.5 (ground truth similarity 0). Cross-entropy loss is utilized to optimize the network. As this method contains several hyperparameters, such as patch sizes and fusion strategies, which can be different for each object to reach best performance, we are still exploring this method to pursue the balance between segmentation accuracy and time efficiency, including embedding part of the idea into the SOMA process. The results presented in Table 6 are based on patches of 15 × 15 pixels with stride 11 inside the ROI determined in the same way as Step R1 of procedure SOMA-R. Different from the original purpose of this method to segment a target object within a cropped region, our focus is more on segmenting all the main organs inside a whole-body region, where the background region may contain different kinds of structures like bone, airway, and soft tissue, and the similarity network needs to be retrained for new objects and it is hard for the network to distinguish among tissues without annotations. The low DC values for TEs (thoracic esophagus) give a good illustration that the local background of TEs is also soft tissue with low contrast, and the similarity network fails to catch the real similarity inside this local region. Besides, we conducted an experiment by applying the similarity network trained for mandible (Mnd) to segment RPG, and the mean DC value degrades from 0.72 to 0.68. Instead, our SOMA recognition method does not only focus on foreground, but also explores the similarity in background region and generates atlas map, which is ready to be used for potential targets. Thus, SOMA can be used in conjunction with any top-of-the-line delineation engine for obtaining the final segmentation.

We would like to summarize the advantages of our proposed SOMA methods as follows: (i) The SOMA method is much less sensitive to the image quality problem arising from artifacts and distortions among real patients. As only the regional best-match atlas is applied to recognize the target object locally, the atlases with lower regional similarity will not influence the recognition quality, especially after selecting the precision subset of atlases for ReR, greatly alleviating this sensitivity problem widely existing in MAS methods.[70] Our collected datasets contain clinical images of 539 cancer patients, 75 of which show overall GQ on all considered objects (meaning that there are no more than three slices with artifacts/deviations, etc.; see[66]) and are used to determine parameters for the recognition process. The SOMA method is evaluated on the rest of the images, composed of 1809 object samples with GQ and 1073 object samples with PQ. As briefly compared in Table 6, recognition and delineation accuracies of SOMA are truly less influenced by object-level image quality variations. (ii) The intermediate recognition results are given in the SOMA process before using DL, which show good localization and refined boundary matching. Recognition alone can be used in clinical analysis, for example, for disease quantification,[71] without having to do delineation. There is rarely such trackable and explainable intermediate result that comes out from DL processes. (iii) The SOMA method conducts subimage matching in both foreground and background regions, which can further extend to the whole-body region and be used in selecting precision atlases for any potential target. This matching process does not explicitly distinguish foreground and background before generating fuzzy membership maps. It can be done only once regardless of the future refinement in atlas

annotations, whereas the DL models need to be retrained when reference masks are updated. (iv) The SOMA method is not specialized to a specific set of objects or body regions, while most of the DL methods aim to segment target objects all at once and are less adjustable when considering other objects, other body regions, or other modalities. The object-specific considerations, including spatial locations, shapes, sizes, and contrasts, are contained in the SOMA recognition process when determining ROIs and parameters θ (similarity threshold) and ω (subimage size), which can be decided separately and will not have influence that permeates among objects.

4.4 | Gaps, challenges, and future works

Several gaps, challenges, and extensions for the SOMA method are to be worked on in the future. First, standardized definitions for body regions and all objects should always be pursued so that the performance of medical image segmentation methods can be reliably and meaningfully compared, although imperfect definitions cannot be entirely eliminated, especially for challenging sparse soft-tissue objects like the esophagus. Such imperfections will lead to errors in localization accuracy of recognition.

Second, the subset of precision atlases is selected by ranking the frequencies of regional best match, although the matching quality is currently only determined by the similarity threshold θ in a binary manner. Instead of a binary decision, there may be potential in combining the frequency with the level of similarity as a new measure to rank the atlases in future work.

Third, since the SOMA method spotlights the selection of precision atlases and the refinement in anatomy recognition, we only utilize a most typical semantic segmentation network—a U-net-based network to transform the fuzzy membership map into the binary segmentation mask. In the literature, some more advanced networks are utilized with different considerations, such as replacement of the ordinary convolution layers by res-block[72,73] or dense-block,[74] or use of a fully convolutional network as the generator network followed by a discriminator network in the generative adversarial network (GAN) strategy.[75,76] To know whether changing the network architecture will yield better delineation accuracy requires further experimentation.

Fourth, according to the division of atlas selection methods defined in,[10,44] the proposed SOMA method is a purely online-learning method where the precision atlases are selected according to the frequencies of best match of intensity-based similarity to the target object in the test image. In our previous work,[38] an atlas grouping method was proposed, and so offline learning and online learning can potentially be combined to improve atlas selection, as well as recognition accuracy and efficiency, in future work.

Finally, we plan to utilize the SOMA method on multimodality image datasets such as positron emission tomography/computed tomography (PET/CT) images. Modifications to be made for use on multimodality cases in terms of how to conduct intermodality registration and the statistics of best-match frequencies are still under investigation. Assessment of the performance of SOMA method in other body regions, such as the abdomen and pelvis, is an additional topic of future research.

4.5 | Computational considerations

The SOMA method was implemented on a computer with the following specifications: 6-core Intel i7–7800X CPU 3.5 GHz with 64 GB RAM, NVIDIA TITAN XP GPU with 12 GB of memory, and GeForce GTX 1070 GPU with 8 GB of memory, running on the Linux operating system. In SOMA, the seven-parameter registration costs 1–2 min for each image. Computational time for recognition depends on the atlas size, subimage size ω, and ROI size. A larger atlas set, ω, and ROI will cost more time in recognition. Recognition of heart (Hrt) in the thorax region is typically the most time consuming, given the large ROI and the largest $\omega = 55$ among all considered objects, which costs around 40 min with 200 atlases. For the mandible (Mnd) with the smallest window size $\omega = 5$, recognition for a test sample costs about 3 min based on 200 atlases. As all objects can be recognized simultaneously, the most time-consuming object determines the recognition time for the entire-body region. In this sense, the recognition of all objects in the thorax region depends on Hrt and is typically 40 min, and the recognition of all objects in the H&N region depends on cervical spinal cord (CtSC), an object with a large spatial extent, with an overall recognition time of about 7 min when using the largest $\omega = 17$. As subimage matching is conducted on each atlas and each position in an ROI, parallel computing is obviously available in the recognition procedure to reduce computational time and will be an area of focus in our future work. DLD costs 1–3 s for each object sample in the test stage, while training time is also based on the number of training samples and the sizes of the ROIs. Typically, training time ranges from ~80 min for small objects like right submandibular gland (RSmG) and up to 10 h for large objects like right lung (RLg).

5 | CONCLUSIONS

In this paper, we introduce a new approach called SOMA of selecting subject-, object-, and modality-adapted precision atlases for automatic anatomy recognition and delineation in medical images with pathology. The proposed method starts from a viewpoint that the recognition of different parts of the target object can be taken from different atlases with best regional (and not global) similarity, while the similarities on other regions do not matter. Hence, the precision atlases have no need to be overall similar to the test image but with frequent regional similarity to the target object, where the frequency of best match is the measure for selecting precise atlases.

The method includes three main components, atlas construction, two-stage recognition, and delineation. The atlas set is constructed by aligning all images into a unified image space, which belongs to the root image determined via an MST strategy from a set of radiologically near-normal images. Then, specific to each test object sample under consideration, RoR is conducted to determine a refined ROI and a set of precision atlases with the highest frequency of regional best match. Subsequently, ReR is conducted with the refined ROI, refined atlases, and a floating window strategy to generate better regional match. A U-net-based deep learning network is trained for delineation, where the original gray scale image together with the fuzzy map from ReR is taken as a two-channel input, while output is the segmentation mask of the target object. We conducted experiments on two body regions,

the H&N region with 298 patient datasets and nine objects, and the thorax with 241 patient datasets and six objects.

We summarize our conclusions as follows. (i) The SOMA method shows high accuracy and robustness in anatomy recognition and delineation. There is a tendency of gradual refinement from RoR to ReR and to delineation, owing to selection of precision atlases in RoR, careful boundary matching in ReR, and strength of deep learning models in interpreting boundaries. (ii) Samples with different object qualities show less difference in recognition and delineation accuracy, whereas the accuracy is obviously influenced by object quality in model-based methods. This confirms one of the SOMA principles that no matter whether an object sample is of good or poor quality, it can be well recognized if there exist partially similar atlases in the atlas set. This is one of the central tenets and strengths of the SOMA approach. (iii) Although only CT images of H&N and thorax body regions are evaluated in the current experiments, the SOMA method is applicable unmodified to other image modalities and other body regions as long as a set of atlases is available such that patterns of different portions of the test sample are able to be represented by a part of the atlas set.

ACKNOWLEDGMENTS

The research reported here is supported by an NIH grant R42CA199735. Jieyu Li's training at the Medical Image Processing Group (MIPG) was supported partly by China Scholarship Council.

Funding information

National Institutes of Health, Grant/Award Number: R42CA199735; China Scholarship Council, Grant/Award Number: 201706230139

DATA AVAILABILITY STATEMENT

Research data are not shared.

REFERENCES

1. Cootes TF, Taylor CJ, Cooper DH. Active shape models-their training and application. Comput Vis Image Underst. 1995;61(1): 38–59.
2. Pizer SM, Fletcher PT, Joshi S, et al. Deformable m-reps for 3D medical image segmentation. Int J Comput Vis. 2003;55(2–3):85–106. [PubMed: 23825898]
3. Shen T, Li H, Huang X. Active volume models for medical image segmentation. IEEE Trans Med Imaging. 2011;30(3):774–791. [PubMed: 21118771]
4. Staib LH, Duncan JS. Boundary finding with parametrically deformable models. IEEE Trans Pattern Anal Mach Intell. 1992;14:1061–1075.
5. Udupa JK, Odhner D, Zhao L, et al. Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images. Med Image Anal. 2014;18(5):752–771. [PubMed: 24835182]
6. Ashburner J, Friston KJ. Computing average shaped tissue probability templates. Neuroimage. 2009;45(2):333–341. [PubMed: 19146961]
7. Christensen GE, Rabbitt RD, Miller MI. 3D brain mapping using a deformable neuroanatomy. Phys Med Biol. 1994;39:609–618. [PubMed: 15551602]

8. Chu C, Oda M, Kitasaka T, et al. Multi-organ segmentation from 3D abdominal CT images using patient-specific weighted-probabilistic atlas. Paper presented at: Medical Imaging 2013: International Society for Optics and Photonics; 2013. 10.1117/12.2007601
9. Gee JC, Reivich M, Bajcsy R. Elastically deforming 3D atlas to match anatomical brain images. J Comput Assist Tomogr. 1993;17:225–236. [PubMed: 8454749]
10. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: a survey. Med Image Anal. 2015;24(1):205–219. [PubMed: 26201875]
11. Shi C, Cheng Y, Wang J, Wang Y, Mori K, Tamura S. Low-rank and sparse decomposition based shape model and probabilistic atlas for automatic pathological organ segmentation. Med Image Anal. 2017;38:30–49. [PubMed: 28279915]
12. Agn M, Af Rosenschöld PM, Puonti O, et al. A modality-adaptive method for segmenting brain tumors and organs-at-risk in radiation therapy planning. Med Image Anal. 2019;54:220–237. [PubMed: 30952038]
13. Cerrolaza JJ, Picazo ML, Humbert L, et al. Computational anatomy for multi-organ analysis in medical imaging: a review. Med Image Anal. 2019;56:44–67. [PubMed: 31181343]
14. Drozdzal M, Chartrand G, Vorontsov E, et al. Learning normalized inputs for iterative estimation in medical image segmentation. Med Image Anal. 2018;44:1–13. 10.1016/j.media.2017.11.005 [PubMed: 29169029]
15. Moeskops P, Viergever MA, Mendrik AM, De Vries LS, Benders MJ, Išgum I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging. 2016;35(5):1252–1261. [PubMed: 27046893]
16. Wang S, He K, Nie D, Zhou S, Gao Y, Shen D. CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation. Med Image Anal. 2019;54:168–178. [PubMed: 30928830]
17. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage. 2009;46(3):726–738. [PubMed: 19245840]
18. Artaechevarria X, Munoz-Barrutia A, Ortiz-de-Solórzano C. Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans Med Imaging. 2009;28(8):1266–1277. [PubMed: 19228554]
19. Asman AJ, Landman BA. Non-local statistical label fusion for multi-atlas segmentation. Med Image Anal. 2013;17(2):194–208. [PubMed: 23265798]
20. Coupé P, Manjón JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. Neuroimage. 2011;54(2):940–954. [PubMed: 20851199]
21. Serag A, Boardman JP, Wilkinson AG, Macnaught G, Semple SI. A sparsity-based atlas selection technique for multiple-atlas segmentation: application to neonatal brain labeling. Paper presented at: 2016 24th Signal Processing and Communication Application Conference (SIU); 2016:2265–2268.
22. van Rikxoort E, Arzhaeva Y, van Ginneken B. A multi-atlas approach to automatic segmentation of the caudate nucleus in MR brain images. Paper presented at: Proceedings of MICCAI 2011 Workshop 3D Segmentation in the Clinic: A Grand Challenge; 2007;29–36.
23. Zhang D, Wu G, Jia H, Shen D. Confidence-guided sequential label fusion for multi-atlas based segmentation. In: Fichtinger G, Martel A, Peters T, eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011. 2011:643–650.
24. Hesterman J, Ghayoor A, Novicki A, et al. Multi-atlas approaches for image segmentation across modality, species and application area. Konica Minolta Technology Report. 2019;16:32–36.
25. Gorthi S, Bach Cuadra M, Schick U, Tercier PA, Allal AS, Thiran JP. Multi-atlas based segmentation of head and neck CT images using active contour framework. Paper presented at: MICCAI workshop on 3D Segmentation Challenge for Clinical Applications; 2010:313–321.
26. Han X, Hoogeman MS, Levendag PC, et al. Atlas-based auto-segmentation of head and neck CT images. Med Image Comput Comput Assist Interv. 2008;434–441. 10.1007/978-3-540-85990-1_52

27. Han X, Hibbard LS, O'Connell NP, Willcut V. Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion. Medical Image Analysis for the Clinic: A Grand Challenge. 2010:297–304.
28. Lee H, Lee E, Kim N, et al. Clinical Evaluation of Commercial Atlas-Based Auto-Segmentation in the Head and Neck Region. Front Oncol. 2019;9:239. [PubMed: 31024843]
29. Raudaschl PF, Zaffino P, Sharp GC, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. Med Phys. 2017;44(5):2020–2036. [PubMed: 28273355]
30. Yang J, Zhang Y, Zhang L, Dong L. Automatic segmentation of parotids from CT scans using multiple atlases. In: Medical Image Analysis for the Clinic: A Grand Challenge. 2010:323–330.
31. van Rikxoort EM, Isgum I, Arzhaeva Y, et al. Adaptive local multiatlas segmentation: application to the heart and the caudate nucleus. Med Image Anal. 2010;14(1):39–49. [PubMed: 19897403]
32. Wolz R, Chu C, Misawa K, Fujiwara M, Mori K, Rueckert D. Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE Trans Med Imaging. 2013;32(9):1723–1730. [PubMed: 23744670]
33. Zhao Y, Li H, Zhou R, Tetteh G, Niethammer M, Menze BH. Automatic multi-atlas segmentation for abdominal images using template construction and robust principal component analysis. Paper presented at: 2018 24th International Conference on Pattern Recognition (ICPR); 2018:3880–3885.
34. Ciardo D, Gerardi MA, Vigorito S, et al. Atlas-based segmentation in breast cancer radiotherapy: evaluation of specific and generic-purpose atlases. Breast. 2017;32:44–52. [PubMed: 28033509]
35. Oliveira B, Queirós S, Morais P, et al. A novel multi-atlas strategy with dense deformation field reconstruction for abdominal and thoracic multi-organ segmentation from computed tomography. Med Image Anal. 2018;45:108–120. [PubMed: 29432979]
36. Schipaanboord B, Boukerroui D, Peressutti D, et al. Can atlas-based auto-segmentation ever be perfect? insights from extreme value theory. IEEE Trans Med Imaging. 2018;38(1):99–106. [PubMed: 30010554]
37. Doan NT, de Xivry JO, Macq B. Effect of inter-subject variation on the accuracy of atlas-based segmentation applied to human brain structures. In: Medical Imaging 2010: Image Processing.. International Society for Optics and Photonics; 2010:76231S.
38. Jin Z, Udupa JK, Torigian DA. How many models/atlases are needed as priors for capturing anatomic population variations? Med Image Anal. 2019;58:101550. [PubMed: 31557632]
39. Lee J, Lyu I, Styner M. Multi-atlas segmentation with particle-based group-wise image registration. Proc SPIE Int Soc Opt Eng. 2014;9034:903447. [PubMed: 25075158]
40. Grevera GJ, Udupa JK, Odhner D, Torigian DA. Optimal atlas construction through hierarchical image registration. In: Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling.. International Society for Optics and Photonics; 2016:97862C.
41. Jia H, Yap PT, Shen D. Iterative multi-atlas-based multi-image segmentation with tree-based registration. Neuroimage. 2012;59(1):422–430. [PubMed: 21807102]
42. Artaechevarria X, Muñoz-Barrutia A, Ortiz-de-Solórzano C. Efficient classifier generation and weighted voting for atlas-based segmentation: two small steps faster and closer to the combination oracle. In: Medical Imaging 2008: Image Processing. International Society for Optics and Photonics; 2008:69141W.
43. Bai W, Shi W, DP O'regan. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. IEEE Trans Med Imaging. 2013;32(7):1302–1315. [PubMed: 23568495]
44. Schipaanboord B, Boukerroui D, Peressutti D, et al. An evaluation of atlas selection methods for atlas-based automatic segmentation in radiotherapy treatment planning. IEEE Trans Med Imaging. 2019;38(11):2654–2664. [PubMed: 30969918]
45. Ramus L, Commowick O, Malandain G. Construction of patient specific atlases from locally most similar anatomical pieces. Med Image Comput Comput Assist Interv. 2010;13:155–162. 10.1007/978-3-642-15711-0_20 [PubMed: 20879395]
46. Rohlfing T, Brandt R, Menzel R. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage. 2004;21(4):1428–1442. [PubMed: 15050568]

47. Katouzian A, Wang H, Conjeti S, et al. Hashing-based atlas ranking and selection for multiple-atlas segmentation. Med Image Comput Comput Assist Interv. 2018;543–551. 10.1007/978-3-030-00937-3_62
48. Langerak TR, van der Heide UA, Kotte AN, Viergever MA, van Vulpen M, Pluim JP. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans Med Imaging. 2010;29(12):2000–2008. [PubMed: 20667809]
49. Wu M, Rosano C, Lopez-Garcia P, Carter CS, Aizenstein HJ. Optimum template selection for atlas-based segmentation. Neuroimage. 2007;34(4):1612–1618. [PubMed: 17188896]
50. Langerak TR, Berendsen FF, Van der Heide UA, Kotte AN, Pluim JP. Multiatlas-based segmentation with preregistration atlas selection. Med Phys. 2013;40(9):91701.
51. Sanroma G, Wu G, Gao Y, Shen D. Learning to rank atlases for multiple-atlas segmentation. IEEE Trans Med Imaging. 2014;33(10):1939–1953. [PubMed: 24893367]
52. Antonelli M, Cardoso MJ, Johnston EW, Appayya MB, Presles B, Modat M. GAS: a genetic atlas selection strategy in multi-atlas segmentation framework. Med Image Anal. 2019;52:97–108. [PubMed: 30476698]
53. Kittler J, Hater M, Duin RP. Combining classifiers. Paper presented at: Proceedings of 13th International Conference on Pattern Recognition; 1996:897–901.
54. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004;23(7):903–921. [PubMed: 15250643]
55. Cardoso MJ, Leung K, Modat M, et al. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. Med Image Anal. 2013;17(6):671–684. [PubMed: 23510558]
56. Vakalopoulou M, Chassagnon G, Bus N, et al. Atlasnet: multi-atlas non-linear deep networks for medical image segmentation. Med Image Comput Comput Assist Interv. 2018;658–666. 10.1007/978-3-030-00937-3_75
57. Zaffino P, Ciardo D, Raudaschl P, et al. Multi atlas based segmentation: should we prefer the best atlas group over the group of best atlases? Phys Med Biol. 2018;63(12):12NT01.
58. Ding Z, Han X, Niethammer M. VoteNet: a deep learning label fusion method for multi-atlas segmentation. Med Image Comput Comput Assist Interv. 2019:202–210.
59. Fang L, Zhang L, Nie D, et al. Brain image labeling using multi-atlas guided 3D fully convolutional networks. Patch Based Tech Med Imaging. 2017;10530:12–19.
60. Huo Y, Xu Z, Xiong Y, et al. 3D whole brain segmentation using spatially localized atlas network tiles. Neuroimage. 2019;194:105–199. [PubMed: 30910724]
61. Yang H, Sun J, Li H, Wang L, Xu Z. Neural multi-atlas label fusion: application to cardiac MR images. Med Image Anal. 2018;49:60–75. [PubMed: 30099151]
62. Li J, Udupa JK, Tong Y, Odhner D, Torigian DA. Anatomy recognition in CT images of head and neck region via precision atlases. In: Medical Imaging 2021: Image Processing. International Society for Optics and Photonics; 2021;1159633.
63. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Med Image Comput Comput Assist Interv. 2015;9351:234–241. 10.1007/978-3-319-24574-4_28
64. Christensen GE, Geng X, Kuhl JG, et al. Introduction to the non-rigid image registration evaluation project (NIREP). Biomed Image Reg. 2006;4057:128–135.
65. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017;1125–1134. 10.1109/CVPR.2017.632
66. Wu X, Udupa JK, Tong Y, et al. AAR-RT – a system for auto-contouring organs at risk on CT images for radiation therapy planning: principles, design, and large-scale evaluation on head-andneck and thoracic cancer cases. Med Image Anal.2019;54:45–62. [PubMed: 30831357]
67. Pednekar GV, Udupa JK, McLaughlin DJ, et al. Image quality and segmentation. Proc SPIE Int Soc Opt Eng. 2018;10576:105762N.

68. Heimann T, Van Ginneken B, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans Med Imaging. 2009;28(8):1251–1265. [PubMed: 19211338]

69. Ding W, Li L, Zhuang X, Huang L. Cross-modality multi-atlas segmentation using deep neural networks. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Vol 12263. Springer; 2020:233–242.

70. Tang H, Chen X, Liu Y, et al. Clinically applicable deep learning framework for organs at risk delineation in CT images. Nat Mach Intell. 2019;1(10):480–491.

71. Tong Y, Udupa JK, Odhner D, Wu C, Schuster SJ, Torigian DA. Disease quantification on PET/CT images without explicit object delineation. Med Image Anal. 2019;51:169–183. [PubMed: 30453165]

72. Chan JW, Kearney V, Haaf S, et al. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. Med Phys. 2019;46(5):2204–2213. [PubMed: 30887523]

73. Zhu W, Huang Y, Zeng L, et al. AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. Med Phys. 2019;46(2):576–589. [PubMed: 30480818]

74. Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Med Image Anal. 2019;51:21–45. [PubMed: 30390512]

75. Dong X, Lei Y, Wang T, et al. Automatic multi-organ segmentation in thorax CT images using U-Net-GAN. Med Phys. 2019;46(5):2157–2168. [PubMed: 30810231]

76. Tong N, Gou S, Yang S, Cao M, Sheng K. Shape constrained fully convolutional DenseNet with adversarial training for multiorgan segmentation on head and neck CT and low-field MR images. Med Phys. 2019;46(6):2669–2682. [PubMed: 31002188]

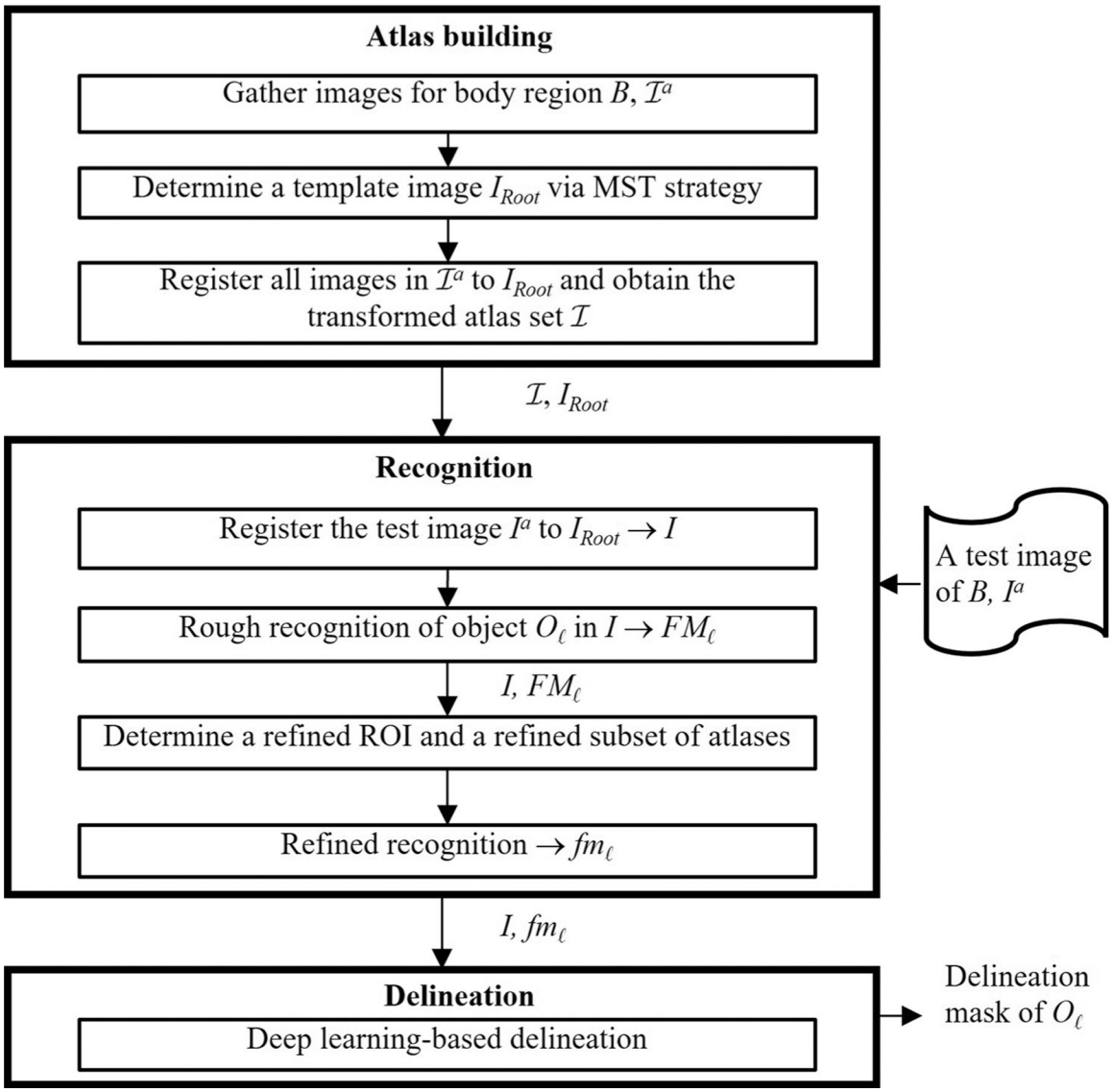

FIGURE 1.
Schematic representation of the SOMA approach

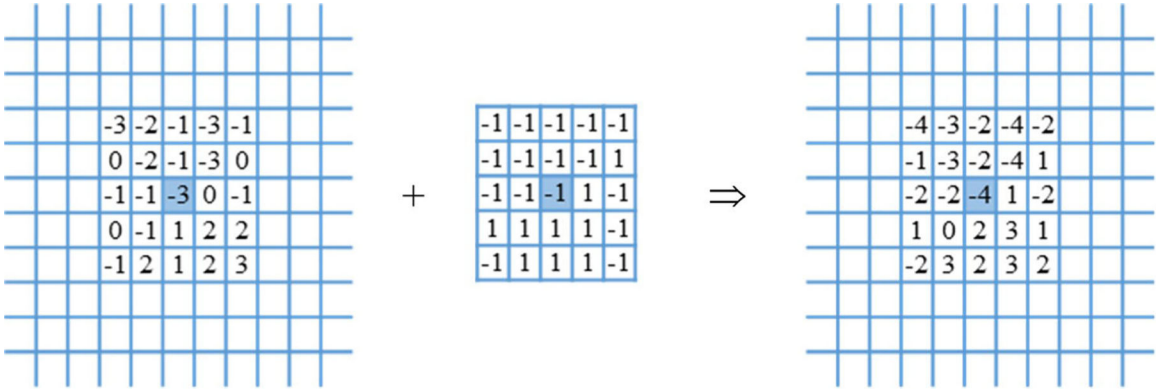

FIGURE 2.

Two-dimensional example of a 5 × 5 subimage $V_{5,,f^*}(v)$ (middle), where v is shown highlighted, the 5 × 5 region around v in FM_f (left), and the same 5 × 5 region around v in the resulting FM_f (right) after the update in Step R6

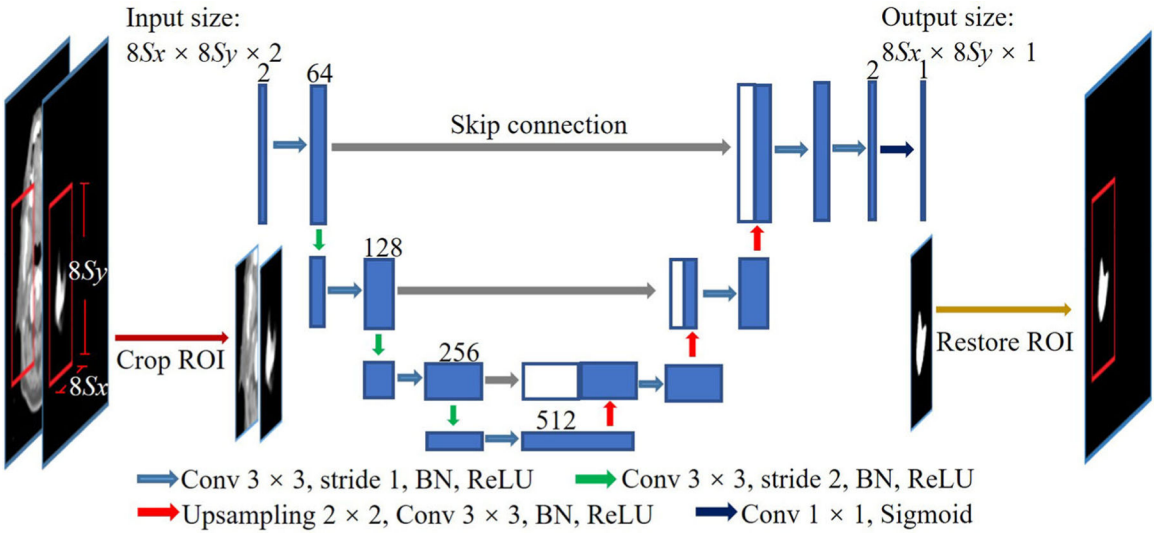

FIGURE 3.

Deep learning network architecture for SOMA delineation procedure. A case of right parotid gland (RPG) is taken as an example

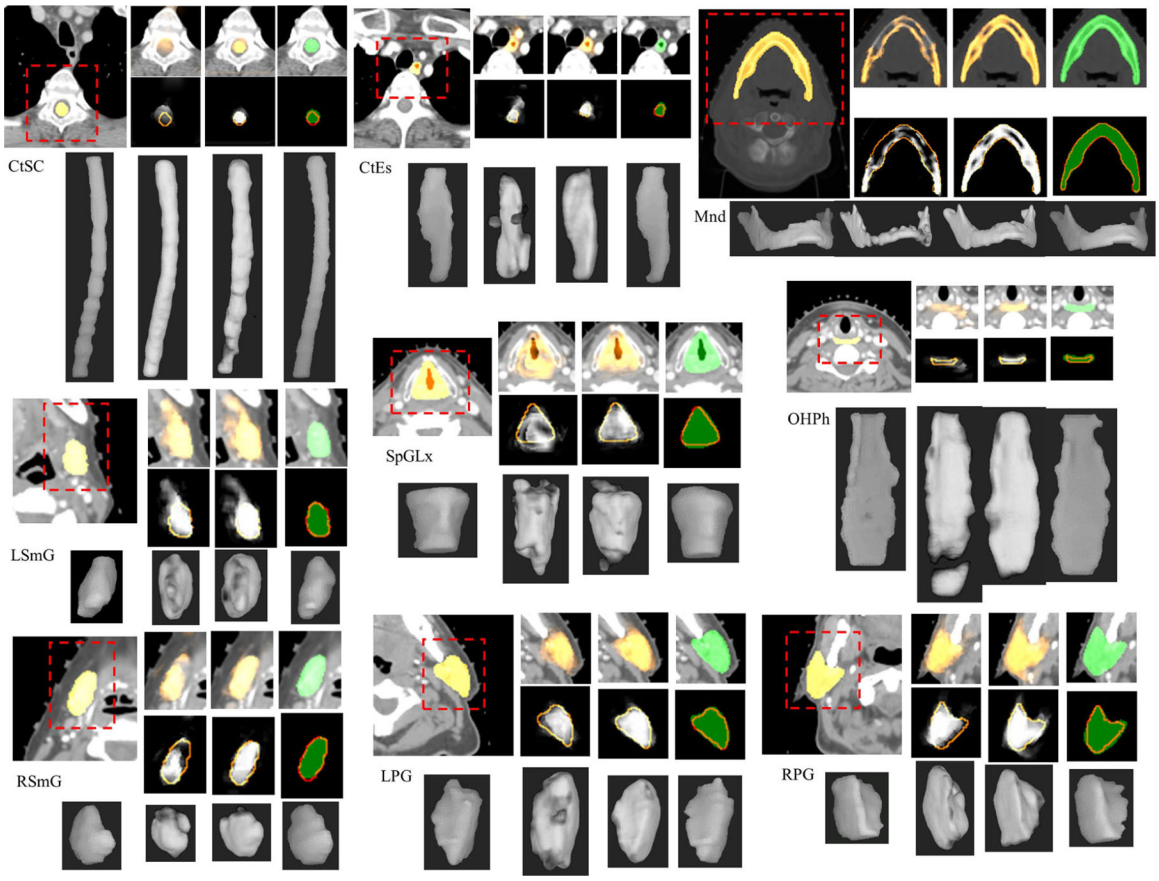

FIGURE 4.

Image examples for objects in the head and neck (H&N) region. Two-dimensional images for reference masks (first column), fuzzy maps from RoR and ReR procedures (second and third columns), and delineation masks (fourth column) overlapped on gray scale images and overlapped by reference contours are shown in first two rows. The corresponding 3D surface or volume renditions are arranged at the bottom

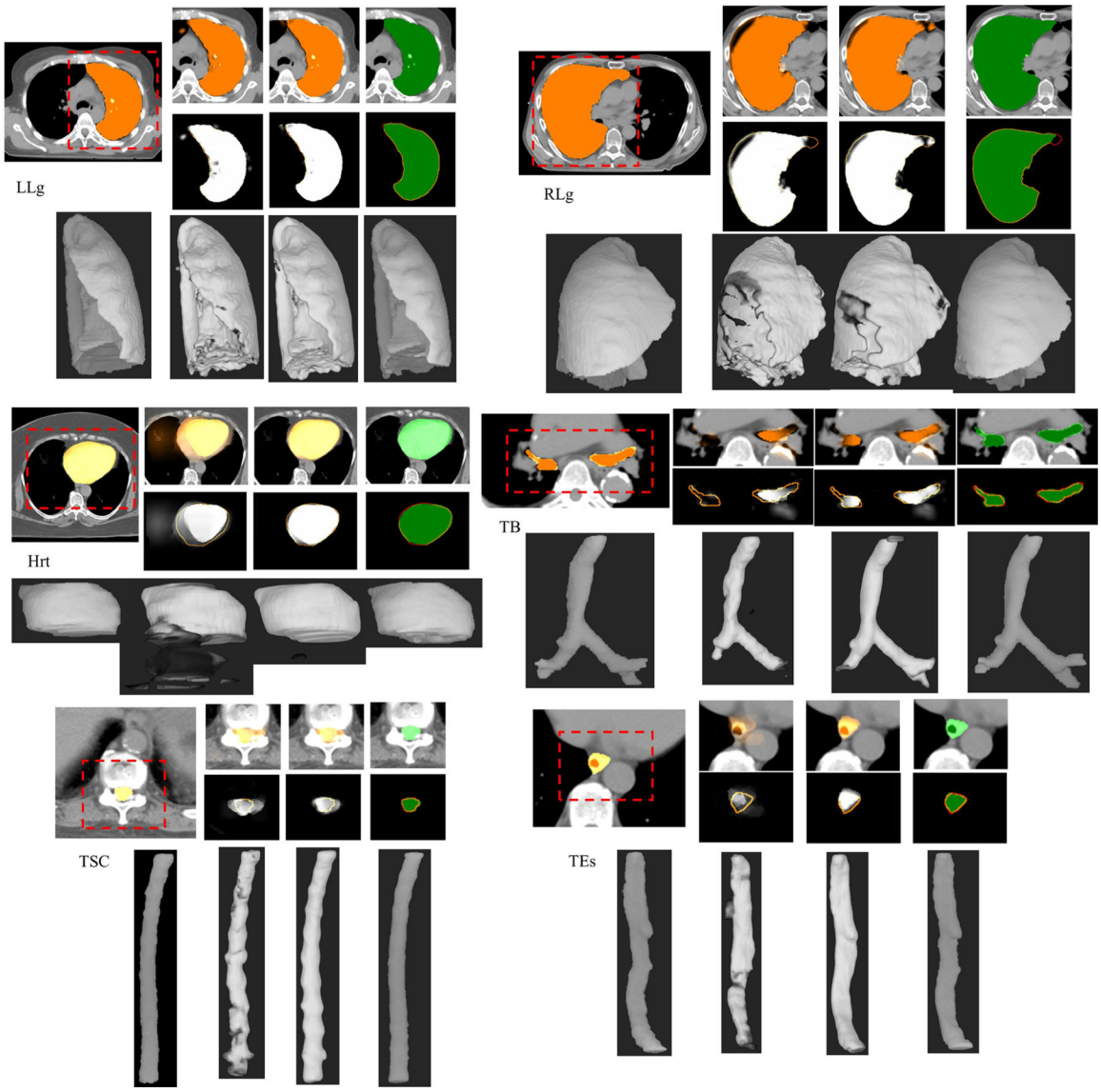

FIGURE 5.

Image examples for objects in the thorax region. Similar to Figure 4, reference masks, recognition maps, and delineation masks are shown from left to right

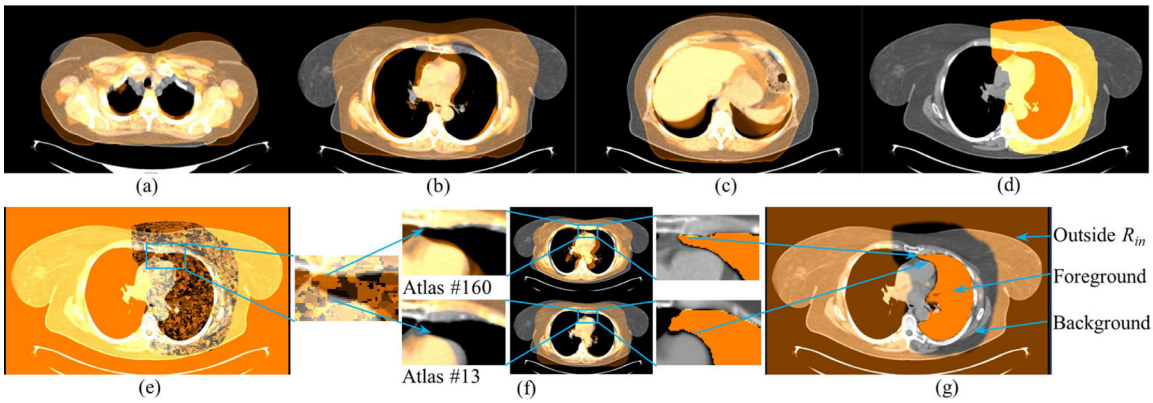

FIGURE 6.

Illustration of why subject-level similarity should not be considered in the recognition process. (a–c) Three slices of the novel image are overlayed by the atlas with best subject-level similarity; this will not guarantee the regional similarity/match. (d) Initial region of interest (R_{in}). (e) Atlas map that indicates indexes of best-match atlases of each region. (f) Different regions of the novel image are matched by different atlases, and their binary masks are combined as in Figure 2 to generate a recognition map as in (g)

TABLE 1

Numbers of samples, folds, and atlases used in experiments

Body region	Objects	Numbers of samples (total/near-normal/GQ/PQ)	N_F	$A_{FR}(\mathcal{J}/\mathcal{J}_R/\mathcal{J}_A)$	T_{eF}
H&N	CtEs (cervical esophagus)	282/36/171/75	3	200/36/164	82
	CtSC (cervical spinal cord)	288/36/169/83	3	200/36/164	84
	Mnd (mandible)	290/36/172/82	3	200/36/164	85
	OHPh (orohypopharynx constrictor muscle)	265/36/51/178	4	200/36/164	58
	LPG (left parotid gland)	211/36/79/96	3	150/36/114	59
	RPG (right parotid gland)	210/36/79/95	3	150/36/114	58
	SpGLx (supraglottic/glottic larynx)	132/36/41/55	3	100/36/64	32
	LSmG (left submandibular gland)	165/36/98/31	2	100/36/64	65
	RSmG (right submandibular gland)	169/36/103/30	2	100/36/64	67
Thorax	LLg (left lung)	240/39/117/84	6	200/39/161	40
	RLg (right lung)	236/39/97/100	6	200/39/161	36
	Hrt (heart)	239/39/146/54	6	200/39/161	39
	TB (trachea and proximal bronchi)	239/39/143/57	6	200/39/161	39
	TSC (thoracic spinal canal)	237/39/186/12	6	200/39/161	37
	TEs (thoracic esophagus)	238/39/157/42	6	200/39/161	38

Abbreviations: A_{FR}, number of atlases contained in $\mathcal{J}$ in each fold; GQ, good quality; H&N, head and neck; N_F, number of folds; PQ, poor quality; T_{eF}, number of test samples in each fold.

TABLE 2

Experimentally determined parameters for subimage size ω and similarity threshold θ for each object

Object	CtEs	CtSC	Mnd	OHPh	LPG, RPG	LSmG, RSmG	SpGLx	LLg, RLg	Hrt	TB	TSC	TEs
ω/θ	9/400	17/200	5/1200	13/400	11/800	7/1200	7/400	7/200	55/200	11/200	17/200	39/200

Med Phys. Author manuscript; available in PMC 2021 December 17.

TABLE 3

Mean (first value) and standard deviation (second value) of quantitative recognition and delineation results of SOMA method in head and neck (H&N) region

			RoR			ReR			DLD	
Object	**OQ**	**Number of samples**	**LE**	**SE**	**DC**	**LE**	**SE**	**DC**	**DC**	**ASD**
CtEs	GQ	171	4.971	0.969	0.628	**3.205**	**0.971**	0.76	**0.818**	0.459
			3.259	0.103	0.119	**2.368**	**0.073**	0.081	**0.059**	0.343
	PQ	75	5.549	0.992	0.616	**3.244**	**0.995**	0.785	**0.805**	0.685
			3.305	0.118	0.157	**2.618**	**0.092**	0.091	**0.146**	1.448
CtSC	GQ	169	5.03	0.983	0.751	**4.205**	**0.994**	0.815	**0.822**	0.553
			4.271	0.047	0.058	**3.287**	**0.028**	0.052	**0.051**	0.34
	PQ	83	6.157	0.97	0.772	**3.12**	**0.993**	**0.853**	0.82	0.6
			5.348	0.066	0.07	**2.713**	**0.026**	**0.063**	0.053	0.319
Mnd	GQ	172	2.098	0.997	0.865	**1.149**	**1**	0.909	**0.924**	0.237
			1.76	0.027	0.065	**0.918**	**0.014**	0.033	**0.021**	0.131
	PQ	82	2.715	0.998	0.85	**1.47**	**0.999**	0.907	**0.921**	0.247
			2.67	0.041	0.088	**1.259**	**0.02**	0.038	**0.028**	0.14
OHPh	GQ	51	6.046	**0.968**	0.553	**5.271**	0.939	0.635	**0.671**	1.018
			4.664	**0.077**	0.107	**4.69**	0.076	0.096	**0.088**	0.578
	PQ	178	4.845	**0.967**	0.580	**4.248**	0.949	0.678	**0.696**	0.864
			3.641	**0.091**	0.108	**2.875**	0.072	0.072	**0.066**	0.397
LPG	GQ	79	3.882	1.018	0.746	**2.626**	**0.99**	0.806	**0.817**	1.092
			2.667	0.089	0.087	**1.577**	**0.081**	0.063	**0.063**	0.512
	PQ	96	3.879	1.013	0.744	**2.534**	**0.993**	0.804	**0.815**	1.078
			2.659	0.102	0.094	**1.558**	**0.086**	0.062	**0.058**	0.53
RPG	GQ	79	3.755	**1.009**	0.742	**2.647**	0.973	0.798	**0.818**	1.044
			2.277	**0.093**	0.081	**1.574**	0.082	0.057	**0.048**	0.409
	PQ	95	3.895	1.013	0.737	**2.804**	**0.992**	0.791	**0.807**	1.2
			2.515	0.106	0.09	**1.993**	**0.097**	0.075	**0.061**	0.654
SpGLx	GQ	41	6.362	1.122	0.649	**3.802**	**1.024**	0.759	**0.783**	3.305
			4.899	0.108	0.13	**3.013**	**0.086**	0.084	**0.095**	2.788

			RoR			ReR			DLD	
Object	**OQ**	**Number of samples**	**LE**	**SE**	**DC**	**LE**	**SE**	**DC**	**DC**	**ASD**
	PQ	55	6.018	1.132	0.627	**4.31**	**1.017**	0.713	**0.747**	4.27
			4.006	0.15	0.104	**3.043**	**0.107**	0.102	**0.115**	4.6
LSmG	GQ	98	4.914	1.125	0.603	**3.339**	**1.062**	0.69	**0.777**	1.061
			3.869	0.208	0.153	**3.659**	**0.231**	0.14	**0.086**	0.818
	PQ	31	6.055	1.127	0.541	**4.881**	**1.084**	0.606	**0.71**	1.215
			3.902	0.158	0.175	**3.686**	**0.164**	0.196	**0.127**	0.899
RSmG	GQ	103	5.001	1.119	0.6	**3.169**	**1.038**	0.685	**0.792**	0.964
			3.993	0.202	0.151	**2.453**	**0.161**	0.136	**0.075**	0.759
	PQ	30	6.142	1.206	0.544	**4.765**	**1.13**	0.616	**0.718**	1.32
			4.559	0.236	0.176	**3.903**	**0.269**	0.171	**0.145**	0.815

Note: Bold numbers show the best results under their corresponding measures.

Abbreviations: ASD, average symmetric distance (in mm); DC, Dice coefficient; DLD, deep learning-based delineation; GQ, good quality; LE, localization error (in mm); OQ, object-level quality; PQ, poor quality; ReR, refined recognition; RoR, rough recognition; SE, scale error.

TABLE 4

Mean (first value) and standard deviation (second value) of quantitative recognition and delineation results of SOMA method in thorax region

			RoR			ReR			DLD	
Object	**OQ**	**Number of samples**	**LE**	**SE**	**DC**	**LE**	**SE**	**DC**	**DC**	**ASD**
LLg	GQ	117	3.001	1.013	0.948	**2.139**	**1.011**	0.96	**0.981**	0.885
			2.407	0.029	0.022	**1.744**	**0.02**	0.014	**0.009**	0.887
	PQ	84	3.907	1.013	0.932	**2.954**	**1.009**	0.946	**0.965**	1.497
			3.945	0.051	0.054	**3.19**	**0.03**	0.047	**0.047**	1.884
RLg	GQ	97	3.149	**1.011**	0.952	**2.961**	1.016	0.96	**0.98**	0.858
			2.808	**0.033**	0.031	**3.235**	0.038	0.027	**0.022**	1.069
	PQ	100	3.993	1.017	0.935	**2.918**	**1.013**	0.951	**0.966**	1.495
			5.446	0.062	0.049	**4.253**	**0.064**	0.04	**0.025**	1.394
Hrt	GQ	146	6.285	1.02	0.865	**4.449**	**1.013**	0.9	**0.92**	2.369
			5.363	0.051	0.083	**4.129**	**0.038**	0.063	**0.045**	1.596
	PQ	54	5.758	1.024	0.871	**3.932**	**1.018**	0.901	**0.915**	3.059
			5.064	0.045	0.08	**3.428**	**0.043**	0.062	**0.049**	3.329
TB	GQ	143	8.729	0.924	0.761	**6.634**	**0.978**	0.832	**0.889**	1.045
			6.715	0.083	0.103	**5.251**	**0.076**	0.059	**0.029**	1.76
	PQ	57	10.039	0.948	0.723	**6.076**	**0.964**	0.821	**0.876**	1.577
			8.531	0.129	0.142	**4.13**	**0.059**	0.082	**0.037**	2.676
TSC	GQ	186	8.613	0.95	0.741	**4.892**	**0.994**	0.845	**0.882**	0.521
			8.601	0.094	0.134	**5.042**	**0.053**	0.061	**0.057**	0.449
	PQ	11	8.589	**0.992**	0.701	**3.672**	1.018	0.82	**0.844**	0.826
			7.019	**0.107**	0.118	**2.26**	0.048	0.065	**0.078**	0.78
TEs	GQ	157	17.393	0.98	0.494	**14.849**	**0.981**	0.612	**0.772**	1.489
			12.273	0.136	0.173	**11.992**	**0.143**	0.151	**0.088**	1.865
	PQ	42	19.055	0.979	0.442	**14.635**	**0.999**	0.55	**0.732**	1.554
			14.644	0.126	0.153	**12.76**	**0.137**	0.151	**0.111**	1.252

Note: Bold numbers show the best results under their corresponding measures.

Abbreviations: ASD, average symmetric distance (in mm); DC, Dice coefficient; DLD, deep learning-based delineation; GQ, good quality; LE, localization error (in mm); OQ, object-level quality; PQ, poor quality; ReR, refined recognition; RoR, rough recognition; SE, scale error.

TABLE 5

Dice coefficient (first value) and localization error (second value) in millimeters for the SOMA recognition results for experiments on empirical parameters with different values for RPG and CtEs

Default: ReR, δ% = 20%, f_r = 2					
RPG	RoR	ReR, δ% = 20%, f_r = 2	ReR_2, δ_2% = 10%, f_r = 2	ReR_3, δ_3% = 5%, f_r = 2	ReR_4, δ_4% = 2.5%, f_r = 2
	0.7381 ± 0.0995	0.7915 ± 0.0747	**0.7919 ± 0.0731**	0.7848 ± 0.0764	0.7726 ± 0.0865
	3.8259 ± 2.8035	2.7344 ± 1.9804	**2.7128 ± 1.9131**	2.8569 ± 2.0116	3.1675 ± 2.2379
	ReR, δ% = 50%, f_r = 2	ReR, δ% = 33%, f_r = 2	ReR, δ% = 20%, f_r = 2	ReR, δ% = 10%, f_r = 2	ReR, δ% = 5%, f_r = 2
	0.7932 ± 0.0739	**0.7933 ± 0.0737**	0.7915 ± 0.0747	0.791 ± 0.0763	0.7788 ± 0.0818
	2.7806 ± 1.9866	**2.7289 ± 1.9692**	2.7344 ± 1.9804	2.8278 ± 1.9599	2.9883 ± 2.0062
	ReR, δ% = 20%, f_r = 0	ReR, δ% = 20%, f_r = 1	ReR, δ% = 20%, f_r = 2	ReR, δ% = 20%, f_r = 3	
	0.7346 ± 0.0964	0.7811 ± 0.079	**0.7915 ± 0.0747**	0.7912 ± 0.0729	
	3.7082 ± 2.7763	3.0148 ± 2.2587	2.7344 ± 1.9804	**2.6489 ± 1.9251**	
CtEs	RoR	ReR, δ% = 20%, f_r = 2	ReR_2, δ_2% = 10%, f_r = 2	ReR_3, δ_3% = 5%, f_r = 2	ReR_4, δ_4% = 2.5%, f_r = 2
	0.6346 ± 0.117	**0.7762 ± 0.0745**	0.7751 ± 0.0755	0.7721 ± 0.0747	0.7635 ± 0.0746
	4.5382 ± 2.9728	2.8728 ± 2.3814	**2.8511 ± 2.277**	2.8746 ± 2.3922	2.945 ± 2.1097
	ReR, δ% = 50%, f_r = 2	ReR, δ% = 33%, f_r = 2	ReR, δ% = 20%, f_r = 2	ReR, δ% = 10%, f_r = 2	ReR, δ% = 5%, f_r = 2
	0.7812 ± 0.0746	0.7804 ± 0.0735	0.7762 ± 0.0745	0.7683 ± 0.0782	0.7527 ± 0.0814
	2.8426 ± 2.4602	**2.8272 ± 2.3688**	2.8728 ± 2.3814	2.9443 ± 2.4094	3.1843 ± 2.3803
	ReR, δ% = 20%, f_r = 0	ReR, δ% = 20%, f_r = 1	ReR, δ% = 20%, f_r = 2	ReR, δ% = 20%, f_r = 3	
	0.6403 ± 0.1012	0.7602 ± 0.0799	**0.7762 ± 0.0745**	0.7738 ± 0.0761	
	4.3034 ± 2.8656	2.9436 ± 2.3616	2.8728 ± 2.3814	**2.8031 ± 2.153**	

Abbreviations: CtEs, cervical esophagus; ReR, refined recognition; RoR, rough recognition; RPG, right parotid gland.

TABLE 6

Comparing mean values of results based on the proposed SOMA method and other methods

	SOMA				AAR-RT[66]				Baseline		Sim_Net	
	LE (mm)		DC (%)		LE (mm)		DC (%)		DC (%)		DC (%)	
Object	**GQ**	**PQ**	**GQ**	**PQ**	**PQ**	**PQ**	**GQ**	**PQ**	**GQ**	**PQ**	**GQ**	**PQ**
CtSC (cervical spinal cord)	4.21	**3.12**	82.2	82	**3.72**	12.47	75.5	55.7	**82.5**	**82.9**	73.2	73.6
Mnd (mandible)	**1.15**	**1.47**	**92.4**	**92.1**	3.58	16.08	87.9	78	92	91.9	81.6	78.8
OHPh (orohypopharynx constrictor muscle)	5.27	**4.25**	67.1	69.6	**3.56**	15.4	52.7	33.6	**67.6**	**70.3**	46.8	48
RPG (right parotid gland)	**2.65**	**2.8**	**81.8**	**80.7**	4.26	11.24	70.7	48.2	80.6	78.7	72	72.1
SpGLx (supraglottic/glottic larynx)	**3.8**	**4.31**	78.3	74.7	3.97	15.99	61.4	39.8	**81**	**75.7**	63.2	62.4
LSmG (left submandibular gland)	**3.34**	**4.88**	**77.7**	**71**	3.39	17.33	70.7	31.8	72.9	61.5	60	54.2
LLg (left lung)	**2.14**	**2.95**	**98.1**	**96.5**	3.7	7.35	95.8	92.5	97.9	94.7	94.1	92.4
Hrt (heart)	4.45	**3.93**	**92**	**91.5**	**4.26**	9.68	84.9	74.6	79.4	77.1	75.8	77
TB (trachea and proximal bronchi)	6.63	**6.08**	**88.9**	**87.6**	**4.13**	8.98	78.8	67.7	85.2	81.9	69.8	64.9
TEs (thoracic esophagus)	14.85	14.64	**77.2**	73.2	**4.88**	**11.92**	55.4	26.4	76.9	**74.2**	14.8	14.9

Note: Bold numbers show the best results under their corresponding measures.

Abbreviations: DC, Dice coefficient; GQ, good quality; LE, localization error; PQ, poor quality.