Automatic upper airway segmentation in static and dynamic MRI via anatomy-guided convolutional neural networks

Xie, L.P.; Udupa, J.K.; Tong, Y.B.; Torigian, D.A.; Huang, Z.H.; Kogan, R.M.; ... ; Arens, R.

Received: 21 May 2021 | Revised: 8 September 2021 | Accepted: 29 October 2021

DOI: 10.1002/mp.15345

RESEARCH ARTICLE

MEDICAL PHYSICS

Automatic upper airway segmentation in static and dynamic MRI via anatomy-guided convolutional neural networks

Lipeng Xie[1,2] | **Jayaram K. Udupa**[2] | **Yubing Tong**[2] | **Drew A. Torigian**[2] | **Zihan Huang**[2] | **Rachel M. Kogan**[2] | **David Wootton**[3] | **Kok R. Choy**[3] | **Sanghun Sin**[4] | **Mark E. Wagshul**[4] | **Raanan Arens**[4]

[1] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China

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

[3] The Cooper Union for the Advancement of Science and Art, New York, New York, USA

[4] Albert Einstein College of Medicine, Bronx, New York, USA

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

Funding information
NIH, Grant/Award Number: HL130468; National Center for Advancing Translational Sciences of the National Institutes of Health, Grant/Award Number: UL1TR001878

Abstract

Purpose: Upper airway segmentation on MR images is a prerequisite step for quantitatively studying the anatomical structure and function of the upper airway and surrounding tissues. However, the complex variability of intensity and shape of anatomical structures and different modes of image acquisition commonly used in this application makes automatic upper airway segmentation challenging. In this paper, we develop and test a comprehensive deep learning-based segmentation system for use on MR images to address this problem.

Materials and Methods: In our study, both static and dynamic MRI data sets are utilized, including 58 axial static 3D MRI studies, 22 mid-retropalatal dynamic 2D MRI studies, 21 mid-retroglossal dynamic 2D MRI studies, 36 mid-sagittal dynamic 2D MRI studies, and 23 isotropic dynamic 3D MRI studies, involving a total of 160 subjects and over 20 000 MRI slices. Samples of static and 2D dynamic MRI data sets were randomly divided into training, validation, and test sets by an approximate ratio of 5:2:3. Considering that the variability of annotation data among 3D dynamic MRIs was greater than for other MRI data sets, we increased the ratio of training data for these data to improve the robustness of the model. We designed a unified framework consisting of the following procedures. For static MRI, a generalized region-of-interest (GROI) strategy is applied to localize the partitions of nasal cavity and other portions of upper airway in axial data sets as two separate subobjects. Subsequently, the two subobjects are segmented by two separate 2D U-Nets. The two segmentation results are combined as the whole upper airway structure. The GROI strategy is also applied to other MRI modes. To minimize false-positive and false-negative rates in the segmentation results, we employed a novel loss function based explicitly on these rates to train the segmentation networks. An inter-reader study is conducted to test the performance of our system in comparison to human variability in ground truth (GT) segmentation of these challenging structures.

Results: The proposed approach yielded mean Dice coefficients of 0.84 ± 0.03, 0.89 ± 0.13, 0.84 ± 0.07, and 0.86 ± 0.05 for static 3D MRI, mid-retropalatal/mid-retroglossal 2D dynamic MRI, mid-sagittal 2D dynamic MRI, and isotropic dynamic 3D MRI, respectively. The quantitative results show excellent agreement with manual delineation results. The inter-reader study results demonstrate that the segmentation performance of our approach is statistically indistinguishable from manual segmentations considering the inter-reader variability in GT.

Conclusions: The proposed method can be utilized for routine upper airway segmentation from static and dynamic MR images with high accuracy and efficiency. The proposed approach has the potential to be employed in other dynamic MRI-related applications, such as lung or heart segmentation.

KEYWORDS
convolutional neural networks, dynamic MRI, image segmentation, static MRI, upper airway

1 | INTRODUCTION

1.1 | Background

Obstructive sleep apnea (OSA) is an increasingly common respiratory disorder, characterized by the narrowing and collapsibility of the pharyngeal airway during sleep.[1] Untreated OSA may cause serious health issues, including cardiovascular disease, metabolic disorders, and cognitive impairment, among others.[2] Common approaches for investigating OSA include polysomnography (PSG) for recording physiological data associated with sleep and breathing,[3] drug-induced sleep endoscopy (DISE) for assessing the airway for anatomical sites of airway narrowing or obstruction,[4] computed tomography (CT) for measuring nasopharynx, oropharynx, and hypopharynx areas,[5,6] and magnetic resonance imaging (MRI) for studying the anatomical structure of the upper airway and surrounding tissues.[7,8] Compared with PSG, DISE, and CT, MRI has the advantages of delineating upper airway anatomy in OSA, locating sites of airway obstruction, guiding surgical treatment,[8] as well as lack of ionizing radiation exposure (for DISE and MRI[9]), affording an increasing popularity in OSA application.

In general, three MR image acquisition modes are commonly employed in OSA—static 3D MRI, dynamic 2D MRI, wherein imaging is acquired rapidly in one slice plane to study upper airway dynamics, and dynamic 3D MRI, wherein full volumetric images are acquired rapidly.[10] In such applications, the delineation of the upper airway in the images is an essential first step for quantitative analysis of the anatomical and functional alterations. Manual contouring is labor-intensive and observer-dependent. In clinical practice, there is an urgent need for automatic, accurate, and efficient upper airway segmentation algorithms. However, development of robust algorithms for this purpose is challenging due to several factors: (1) complex variability of grayscale intensity among different MR image acquisition protocols, (2) shape variability of upper airway across different patients and slice orientations, (3) low image contrast between upper airway and surrounding tissues, and (4) often inadequate signal-to-noise ratio.

1.2 | Related work

Liu et al.[11] designed a computerized system based on a fuzzy connectedness algorithm[12] to segment the upper airway from axial T2-weigted MR images and tested the software package on 40 patients, reporting a high accuracy rate of 0.97 and an average processing time per case of about 4 min. A minimally interactive model for 4D dynamic MRIs based on an iterative relative fuzzy connectedness delineation algorithm was proposed in Ref. 13, where seeds need to be specified in the 3D images of only the first time instance of the whole 4D dynamic MRI. Yet, this method is affected by noise and poor image contrast. Inspired by the powerful feature learning ability of deep learning (DL) techniques,[14] we employed DL-based models to fully automatically delineate the upper airway from 4D MRIs. In Ref. 15, we constructed a modified 2D U-Net[16] by making full use of the spatial and temporal information from the MR images achieving a Dice cofficient of 0.88. However, this methdology was applicable only to 4D MRIs.

In this study, we present a unified framework to effectively handle upper airway segmentation from all MR image acquisition approaches, whether static or dynamic, or 2D or 3D. Several other machine learning-based upper airway segmentation approaches have been proposed in recent years. Ivanovska et al.[17] developed a semiautomatic method to segment the pharyngeal structures in MRI, in which the initialization step requires users to provide the locations of the starting slice and the parapharyngeal fat pads. Javed et al.[18] utilized 3D region growing to construct a pipeline for delineating the upper airway structure from dynamic 3D MRI, which requires the user to locate the region-of-interest (ROI) in the image and place two seeds in the upper airway. Shahid et al.[19] proposed an automatic method to segment the pharynx from MRI by using traditional image segmentation methods and machine learning models. To study the mechanism of speech production, Silva and Teixeira[20] presented an unsupervised segmentation model to segment the vocal tract airway from real-time MRI based on an active appearance model. In Ref. 21, Li et al. built a set of linked statistical shapes by principal component analysis, then localized the upper airway region by a hierarchical model-fitting algorithm,

and finally segmented the head and neck upper airway by a multifeature-level set contour propagation scheme.

Similarly, several DL-based upper airway segmentation approaches were proposed. For instance, Erattakulangara and Lingala[22] utilized U-Net to segment the vocal tract airway from mid-sagittal 2D MRI and Liu et al.[23] developed a synthetic MR-aided dual pyramid network to segment multiorgan of head and neck from CT images, including pharynx structure. However, the above approaches have limitations: (1) limited efficiency: semiautomatic methods often require much manual intervention during the segmentation procedure, which becomes impractical when dealing with dynamic MRI data sets acquired over several respiratory cycles that can include hundreds (in 2D dynamic MRI) or even thousands (in 3D dynamic MRI) of slices in one study. (2) Inadequate description and usage of anatomical knowledge: the above approaches did not provide precise definitions of the upper airway body region and did not apply anatomical knowledge in the model construction stage. (3) Inadequate application scope: the above approaches can segment the upper airway in only one type of specific MRI protocol without demonstrated generalizability.

A general system that can segment images from any of the three modes of MR image acquisition in this application area in a production mode is currently lacking. In this paper, we propose a highly automated general system based on an *anatomy-guided* deep convolutional neural network (CNN) that operates on all three modes and produces results that are in high agreement with reference standard manual segmentations. The main innovations and contributions of our study are as follows: (1) a complete and unique system for automatic and production-mode segmentation of upper airway structures in different forms of static and dynamic MRI commonly employed in the study of OSA. (2) A precise anatomical definition of upper airway body region in static and dynamic MRI for OSA application. (3) A judicious design of a methodology combining human-knowledge and DL techniques to tackle a very challenging segmentation problem. (4) A loss function based on segmentation false positives (FPs) and false negatives (FNs) that can force the DL model to focus on the explicit reduction of FPs and FNs during the training stage. (5) Demonstration of close to the highest possible performance for the OSA application considering the quality of images and the variability in ground truth (GT). (6) Comparison with other methods from the literature that can reflect the comprehensive advantages of our study for OSA application.

The details about our approach are introduced in Section 2, including data acquisition, anatomy-guided DL principle, and segmentation network design. In Section 3, we demonstrate results, including metric values for quantitatively analyzing the segmentation performance, and comparisons between our approach and related studies. We summarize our conclusions in Section 4.

A preliminary report on this work appeared in the proceedings of the 2021 SPIE Medical Imaging Conference.[24] The present paper includes the following enhancements over the conference paper: (1) a detailed and comprehensive literature review about upper airway segmentation in MRI. (2) A full description of our approach, including the definition of the upper airway in three modes of MR image acquisition, a generalized region-of-interest (GROI) strategy, and anatomy-guided segmentation models. (3) A full summary of our experimental results, including illustrative segmentation examples, quantitative evaluation metric values, inter-reader study results, and comparisons with other approaches in the literature.

2 | MATERIALS AND METHODS

2.1 | Method overview

Figure 1 is a schematic depiction of the main stages in our study. In the preparatory stage, we acquired and annotated static 3D, dynamic 2D, and isotropic dynamic 3D MRI scans. Each group contains four parts: data sets for training, validation, testing, and inter-reader variability study for evaluating the performance of our system. The model-building stage involves the following steps: (1) defining precisely the anatomical structure of the upper airway of focus for our study for the above three modes of MR image acquisition and dividing the upper airway in static 3D MRI into two subregions due to the great shape variability among different portions of the upper airway; (2) developing a GROI strategy for localizing (recognizing) the upper airway based on the statistics of the geometric layout of annotated data sets; (3) designing two DL-based segmentation networks (models) for static and dynamic MRI guided by the anatomical knowledge of the upper airway; (4) constructing the network based on U-Net and contextual information; (5) devising a new loss function based on FN and FP concepts; and (6) performing data augmentation and DL model training. In the evaluation stage, we employed the trained models to segment MRIs in the testing and inter-reader sets.

2.2 | Preparatory operations

2.2.1 | Data acquisition

All data sets in this study were acquired on a Philips 3.0T MRI scanner, under an ongoing prospective IRB-approved research study protocol at the Albert Einstein College of Medicine. The deidentified image data sets

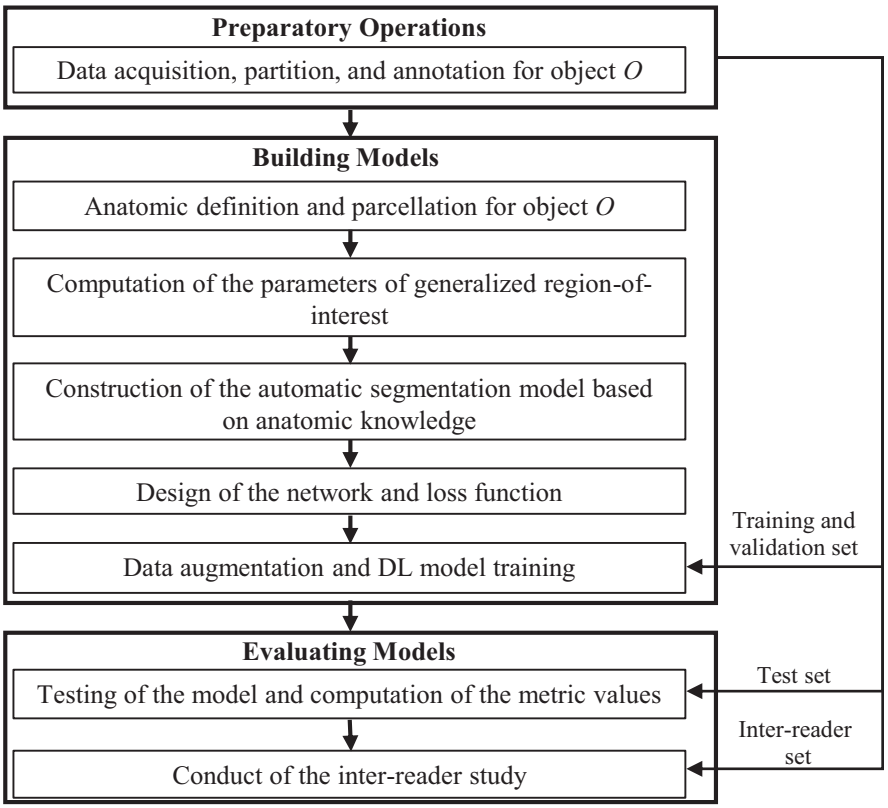

FIGURE 1 Processing pipeline depicting the main stages of our study

TABLE 1 Summary of available upper airway MRI studies under various image acquisition protocols

Data type		3D Voxel size (mm^3)/2D pixel size (mm^2)	Training	Validation	Testing	Total subjects/frames
Static 3D MRI (T2 weighted)		0.50 × 0.50 × 3.30	30	6	22	58 (2133 slices)
Dynamic 2D MRI (balanced SSFP, T1/T2 weighted)	mid-retropalatal	0.94 × 0.94	12	4	6	22 (1979 slices)
	mid-retroglossal	0.94 × 0.94	11	4	6	21 (1962 slices)
	mid-sagittal	0.94 × 0.94	20	6	10	36 (3440 slices)
Isotropic dynamic 3D MRI (gradient echo, T2* weighted)		1.50 × 1.50 ×1.50	10 subjects (164 3D volumes with 7544 slices)	3 subjects (25 3D volumes with 1150 slices)	10 subjects (40 3D volumes with 1840 slices)	23 subjects (229 3D volumes with 10 534 slices)

were subsequently sent to and analyzed at the University of Pennsylvania.

The MRI sequences utilized are summarized in Table 1. Further details about MR image acquisition are as follows:

(i) Static MRI scans were from 58 subjects in the wake condition (34 males, 24 females, age range 8–18 years, and mean age 14±2 years), including 29 with OSA, using parameters: TE/TR = 80/3940 ms, echo train length = 29, voxel size = 0.50 × 0.50 × 3.30 mm^3, FOV = 240 × 240 mm^2, SENSE factor = 2.0, and acquisition time = 1.5 min. The number of axial slices in each study was in the range 36–60.

(ii) Dynamic 2D MRI scans were at three locations: (1) slice plane at mid-retropalatal region, which is orthogonal to the airway centerline (11 subjects, 5 females, age range 12–17 years, mean age 15±2 years, and 5 with OSA), (2) slice at mid-retroglossal region (11 subjects, 5 females, age range 12–17 years, mean age 15±2 years, and 5 with OSA), and (3) slice at mid-sagittal region (18 subjects, 8 females, age range 12–18 years, mean age 15±2 years, and 6 with OSA). In total,

this data set contained 22 mid-retropalatal dynamic 2D MRI sequences, 21 mid-retroglossal dynamic 2D MRI sequences, and 36 mid-sagittal dynamic 2D MRI sequences, which were acquired during sleep and wake states. Considering that the discrepancy of upper airway size, shape, and location between dynamic 2D MRI sequences of the mid-retropalatal and mid-retroglossal is minor, we combined these two dynamic 2D MRI sequences into one data set for segmentation purposes. The number of time instances was 300 in each 2D dynamic sequence. Three thousand nine-hundred and forty-one slices of mid-retropalatal and mid-retroglossal sequences, and 3440 slices of mid-sagittal sequences were manually delineated. The parameters for mid-sagittal acquisitions were as follows: TE/TR = 1.56/3.1 ms, flip angle = 35°, single slice, FOV = 180 × 180 mm^2, temporal resolution = 330 ms/frame, SENSE factor = 1.2, and acquisition time = 2.4 min. The parameters for mid-retropalatal (or mid-retroglossal) acquisitions were as follows: TE/TR = 2.5/5.0 ms, flip angle = 35°, single slice, FOV = 180 × 180 mm^2, temporal resolution = 330 ms/frame, SENSE factor = 1.2, and acquisition time = 2.4 min.

(iii) Isotropic dynamic 3D MRI scans were from 23 completely different patients/controls (16 males, 7 females, age range 12–18 years, and mean age 15±2 years), including 16 with OSA, where 480 frames (3D volumes) were acquired during the sleep state and 480 frames (3D volumes) were acquired during the wake state in each subject, and where each 3D volume contained a total of 46 sagittal reconstructed slices. Images in 229 frames (160 in sleep state and 69 in wake state) in this data set were manually delineated. The isotropic dynamic 3D MRI protocol was as follows: TE/TR = 2.1/4.8 ms, FOV = 200 × 170 mm^2, y-z radial k-space traversal with golden-angle sampling, and L1-ESPIRiT, compressed sensing reconstruction, acquisition time = 4 min.

The data sets used for training, validation, and testing of the DL network are summarized in Table 1. Since our DL method performs segmentation on 2D slices, we also display the number of slices in each type of data set. Overall, this study included a total of 20 048 image slices.

2.2.2 | Data annotation

Data annotation (manual segmentation) is one of the essential procedures required for training a DL model. Since public data sets for upper airway segmentation are currently unavailable, our study team constructed a sufficiently large database of manual segmentations for each of the three modes of MR image acquisition by having trained members of our team perform data annotations that were then checked by a radiologist with over 20 years of experience in MRI. The open-source software CAVASS[25] was adopted for delineation. In addition, we conducted an inter-reader study where we randomly selected a subset of annotated MRI scans for each mode of image acquisition, and had a different trained team member to independently annotate these data sets. In total, 2785 slices with two sets of data annotation were included in the inter-reader study, as shown in Table 2. The purpose of the inter-reader study was to determine the variability in GT delineation and how the performance of our system compared with this variability.

TABLE 2 Summary of the available upper airway MRIs for the inter-reader study

Data type	Subjects
Static 3D MRI	10 (997 slices)
Mid-retropalatal and mid-retroglossal dynamic 2D MRI	10 (503 slices)
Mid-sagittal dynamic 2D MRI	10 (733 slices)
Isotropic dynamic 3D MRI	3 (12 frames with 552 slices)

2.3 | Building models: anatomy-guided DL

There are multiple anatomical structures in the acquired MR images, including the upper airway, paranasal sinuses, skeleton, tonsils, tongue, muscles, and so on, where similarities in intensity distribution among them might lead to FP segmentation results. Moreover, the lack of a precise body region definition will result in great variability among the manual annotation data. Therefore, to enhance the specificity and effectiveness of our DL networks in object delineation, following previously proposed body-wide automatic anatomy recognition (AAR) principles,[26] we provide object recognition help guided by human anatomical knowledge so that the DL networks will always operate in the consistently specified region of the localized anatomy.

This knowledge is brought to bear in three stages: (1) use of a standardized object definition; (2) anatomical parcellation; and (3) use of a GROI. Based on anatomical knowledge, we utilize DL networks to construct the upper airway segmentation models for the three modes of MR image acquisition.

2.3.1 | Anatomical definition

We define a standardized body region definition for each of the three modes of MR image acquisition, in order to

TABLE 3 Summary of body region definition of upper airway for the three modes of MRI

Data type	Body region definition of upper airway		
	Superior boundary	**Inferior boundary**	**Components of upper airway**
Static 3D MRI	Superior aspect of the globes	Level of the true vocal cords	Nasal cavity, lumen of nasopharynx, oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway
Mid-retropalatal dynamic 2D MRI	At mid-retropalatal region in a plane orthogonal to nasopharynx airway centerline		
Mid-retroglossal dynamic 2D MRI	At mid-retroglossal region		
Mid-sagittal dynamic 2D MRI	Superior most aspect of the hard palate	Level of the true vocal cords	Oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal lumen
Isotropic dynamic 3D MRI	Inferior aspect of the sphenoid sinus	Level of the true vocal cords	Lumen of nasopharynx, oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway

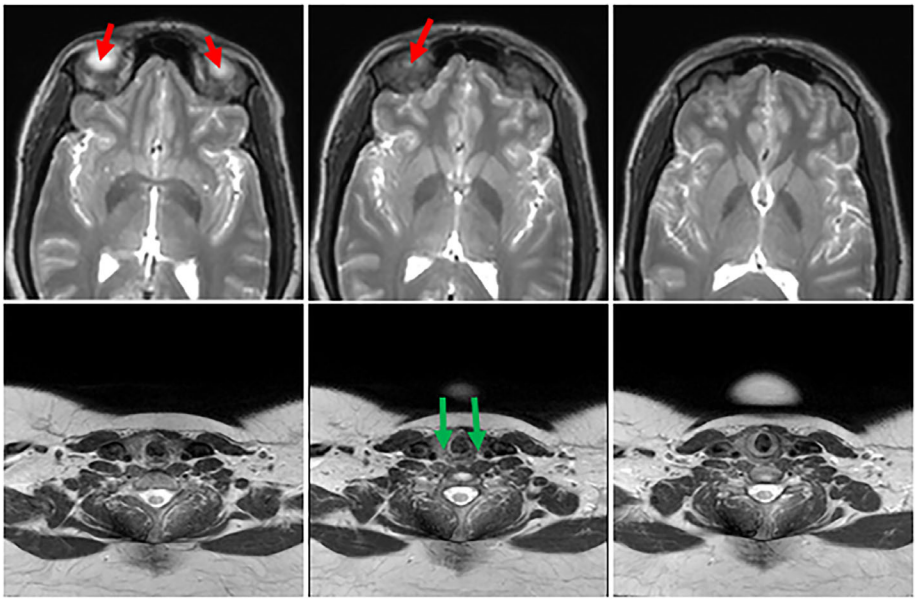

FIGURE 2 Illustration of the boundaries of the upper airway in static 3D MRI. The middle slices in top and bottom rows show the superior aspect of the globes (arrows) and the level of the true vocal cords (arrows), respectively. The slices on the left and right are immediately inferior and superior to the boundary slices in the middle

locate the upper airway precisely and remove adjacent tissue regions, as described in Table 3. Note that the region of focus of the upper airway is not necessarily the same in these different modes of imaging. Accordingly, the superior and inferior boundaries of the structure of focus are defined. For static 3D MRI, we define the superior aspect of the globes (whichever is most superior) as the superior boundary of the body region of interest, and the inferior boundary to be the level of the true vocal cords, as shown in Figure 2. For mid-sagittal dynamic 2D MRI, we define the superior most aspect of the hard palate and the level of true vocal cords as the superior and inferior boundaries of the upper airway, respectively, as shown in Figure 3a. The mid-retropalatal and mid-retroglossal dynamic 2D MR images were acquired from the mid-retropalatal and mid-retroglossal regions during dynamic image acquisition, as illustrated in Figure 3b. For isotropic dynamic 3D MRI, we define the inferior aspect of the sphenoid sinus and the level of the true vocal cords as the superior and inferior boundaries of upper airway, respectively, as shown in Figure 3c. Note that there is slight discrepancy among the three modes of MR image acquisition in terms of the components of the upper airway covered in the definitions.

2.3.2 | Anatomical parcellation

In static 3D MRI, the upper airway is composed of the nasal cavity, lumen of nasopharynx, oropharyngeal,

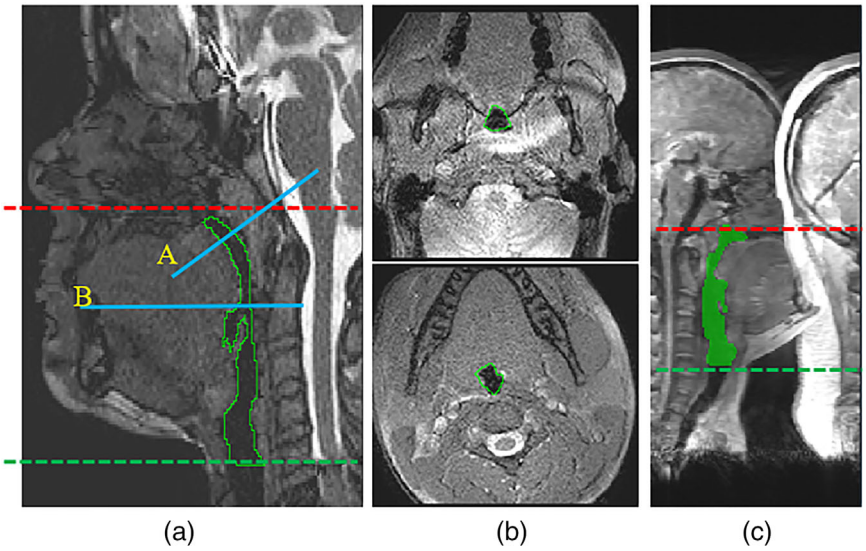

FIGURE 3 Illustration of the upper airway in (a) mid-sagittal dynamic 2D MRI, (b) mid-retropalatal region and mid-retroglossal dynamic 2D MRI, and (c) isotropic dynamic 3D MRI. The red and green lines represent the superior and inferior boundaries, respectively. Blue solid lines labeled A and B represent the orthogonal retropalatal and axial retroglossal levels, respectively. The upper airway boundaries are delineated in green

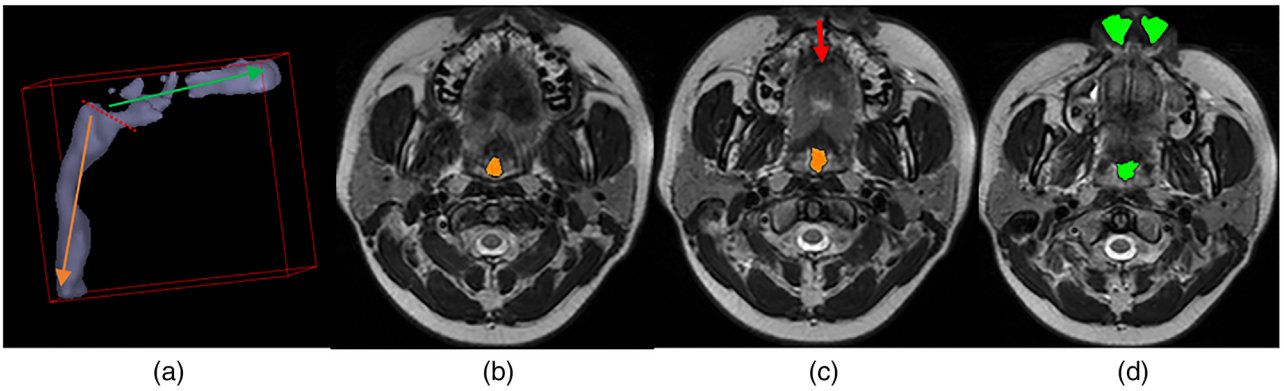

FIGURE 4 Illustration of anatomic parcellation of upper airway in static 3D MRI: (a) representative example of upper airway 3D surface rendition (green arrow: the nasal cavity and lumen of nasopharynx, orange arrow: oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway, and red dotted line: boundary between two subobjects); (c) boundary slice between the two subobjects (red solid arrow: the inferior aspect of the hard palate); (b) and (d) slices immediately inferior and superior to the boundary slice. The orange and green colors represent the oropharyngeal airway and nasopharyngeal airway, respectively

hypopharyngeal, and supraglottic/glottic laryngeal airway. As shown in the 3D surface rendition in Figure 4a, the upper airway is a tubular structure, where the air passages in the nasal cavity are approximately at right angles with respect to the other portions of the pharyngeal and laryngeal airway. As such, in axial images, the air passages in the nasal cavity are roughly in the plane of acquisition, while those in the rest of the airway are generally orthogonal to the scan plane. Therefore, the variability in shape and appearance is significant among different portions of upper airway in static 3D MRI. This makes segmentation of the nasal cavity much more challenging than for the rest of the airways. We, therefore, divide the upper airway in static 3D MRI into two subobjects: an upper portion, including the nasal cavity and nasopharyngeal airway, and a lower portion, including the oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway. The inferior axial boundary of the first subobject is defined to be the inferior aspect of the hard palate as shown in Figure 4c. The anatomical parcellation was not needed and hence not applied in other modes of MR image acquisition since the challenge for segmentation was not so markedly different in different parts of the object being segmented.

2.3.3 | Generalized region-of-interest

Since the position of the upper airway is relatively fixed, we exploit the spatial information of the annotated MRI data set and devise a GROI strategy for recognizing the main object of interest and reducing the negative effect of the tissues located far away from the upper airway in the entire image. Since our segmentation strategy is slice-by-slice, the GROI process described below

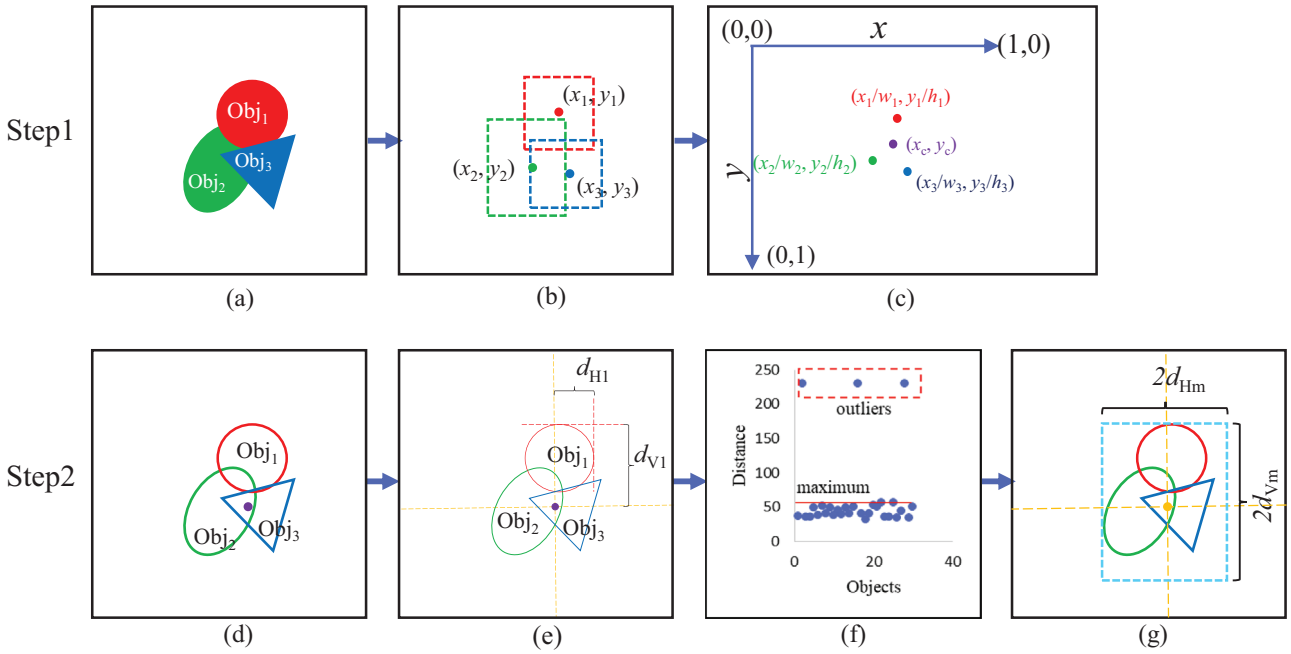

FIGURE 5 Illustration of the GROI strategy: (a) representative examples of annotated objects, (b) bounding boxes of objects, (c) normalized center points, (d) boundaries of objects, (e) maximum distance between center points and object boundary, (f) scatter plot of distance values, and (g) window size of GROI

is applied to each 2D image (slice) in all modes of data acquisition.

The training stage of the GROI strategy involves two steps: (1) locating the center point of the ROI, and (2) determining the window size of the ROI. Assume that the annotated 2D MR slice I_i with size $w_i \times h_i$ involves an upper airway object Obj_j. In the first step, we determine a fitting bounding box for Obj_j to estimate the center position (x_i, y_i) of Obj_j and normalize the horizontal and vertical distance of the center position to the range [0, 1] by dividing the distance by the size of the original image: $(x_i/w_i, y_i/h_i)$, as shown in Figure 5a–c. In the second step, we calculate the mean value of center positions of all samples of Obj_j as the center point of ROI: $\{(x_c, y_c) \mid x_c = \frac{1}{N}\sum_{i=1}^{N}\frac{x_i}{w_i},\ y_c = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{h_i}\}$, where N is the number of annotated samples of Obj_j. Then, we place the center point in the original image I_i with the coordinate values of $(w_i x_c, h_i y_c)$ and measure the maximum distance between the center point and the bounding box boundary of Obj_j in the horizontal and vertical directions (d_{Hi}, d_{Vi}), as illustrated in Figures 5d and e. To refine the size of the ROI, we utilize a scatter plot to analyze the distribution of these x and y distance values taken together and remove the outlier samples, as illustrated in Figure 5f. The outliers are often generated by inconsistent annotation data. Finally, we calculate the horizontal and vertical maximum distance (d_{Hm}, d_{Vm}) among the remaining samples, and utilize $2d_{Hm} \times 2d_{Vm}$ as the window size of the ROI, as illustrated in Figure 5g. Considering that the max-pooling operations with stride 2 in the proposed network narrow the resolution of input ROI to 1/16 of the original size, we make a small modification of $(2d_{Hm}, 2d_{Vm})$ to ensure that $(2d_{Hm}, 2d_{Vm})$ are multiples of 16, eliminating the rounding error.

In the application stage of the GROI strategy, the center position of the ROI is varied with the image size, while the ROI size is fixed and directly determines the input data size for the proposed network. Given an MRI slice with image size $w \times h$, we utilize (wx_c, hy_c) and $2d_{Hm} \times 2d_{Vm}$ as the center position and window size of ROI, respectively, to draw an ROI in the MR image slice automatically, which covers the upper airway adequately in all three modes of MR image acquisition, as illustrated in Figure 6.

2.3.4 | Building DL models

As mentioned in Ref. 27, the same tissue in MR images shows significant variability of intensity across patients and scanners, leading to the poor segmentation performance for automatic methods. To handle this problem, we employed the intensity standardization approach[28] to preprocess all acquired MR images, wherein the original intensity scale is transformed to a standardized scale.

We build two different DL models for static and dynamic MRI as depicted in Figure 7. The static MRI DL model (Figure 7a) consists of two pipelines—one

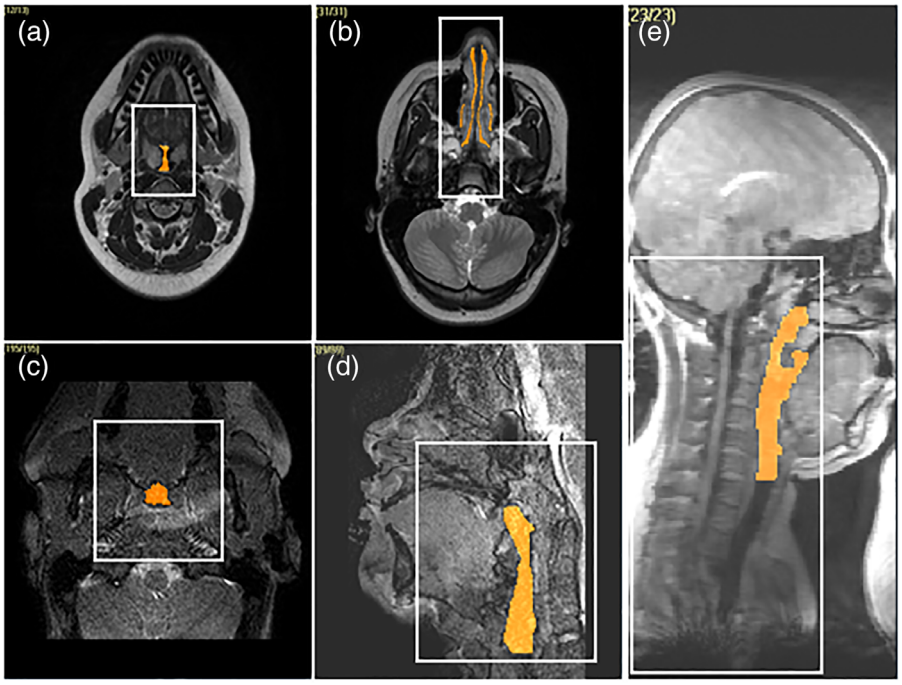

FIGURE 6 Examples of automatic ROI placement in MR images: (a), (b) pharyngeal/laryngeal airway and nasal cavity in static 3D MRI; (c), (d) oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal lumen in dynamic 2D MRI; and (e) upper airway in isotropic dynamic 3D MRI

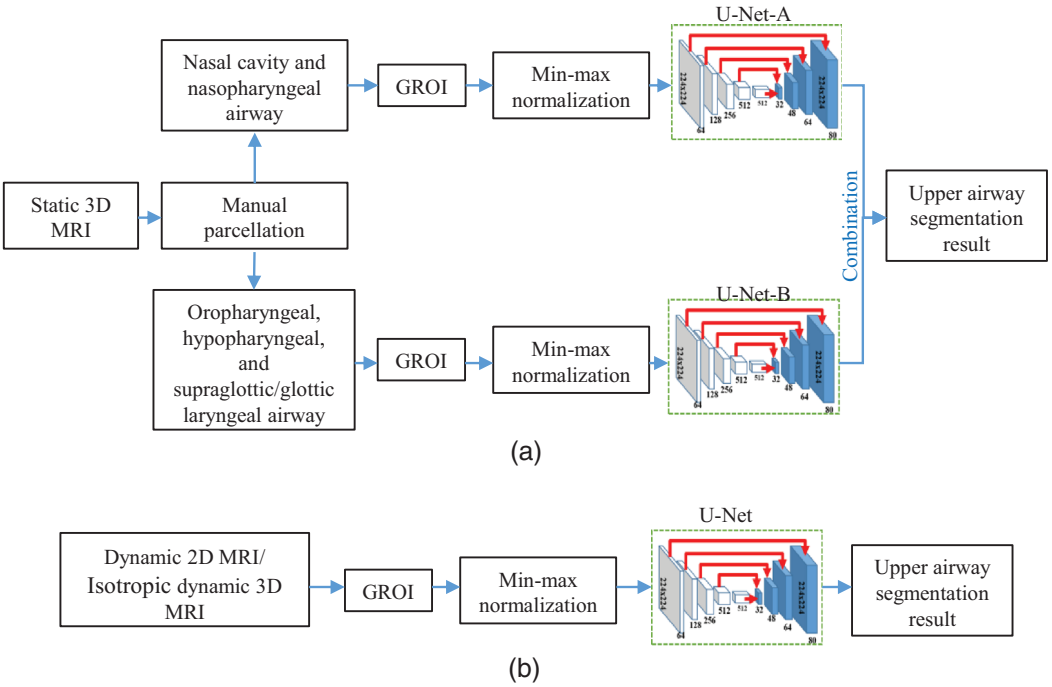

FIGURE 7 Illustration of anatomy-guided DL segmentation models for: (a) static MRI and (b) dynamic MRI

for segmenting the nasal cavity and nasopharyngeal airway, and another for oropharyngeal, hypopharyngeal, and supraglottic/glottic airway. First, the user localizes the superior slice of nasal cavity and nasopharyngeal airway, the inferior slice of the hard plate, and the oropharyngeal, hypopharyngeal, and supraglottic/glottic airway, which leads to two separated 3D subvolumes as inputs for specific DL models. According to this

anatomical information, we intercept these two separate portions of the upper airway from the static 3D MR images and input them into the two pipelines. Then, the proposed GROI strategies with learned parameters are employed to localize the centers of the ROIs in the slices for these two upper airway subvolumes. We extract the ROI data from the slices by image cropping and utilize the min–max normalization method to preprocess the ROI data, rescaling the range of intensity to [0, 1]. The min–max normalization is defined as:

$$I_{\text{norm}}(x) = \frac{I(x) - I_{\min}}{I_{\max} - I_{\min}}, \quad (1)$$

where $I(x)$ denotes the standardized intensity value at pixel x, and $I_{\min}$ and $I_{\max}$ represent the minimum and maximum standardized intensity value among all pixels in image I. Subsequently, the DL network extracts the hierarchical features from the normalized ROI data and classifies each pixel in ROI into two categories: upper airway with label "1" and background with label "0." To keep the size of outputs the same as the original input MR images, we utilize zero-padding to embed the segmentations in the full image. Finally, we combine the segmentations of the two pipelines together as the complete upper airway segmentation result of the 3D volume.

A similar approach is applied to segment other MRI scans acquired using mid-retropalatal dynamic 2D MRI, mid-retroglossal dynamic 2D MRI, mid-sagittal dynamic 2D MRI, and isotropic dynamic 3D MRI with the same pipeline structure, as shown in Figure 7b. However, for these modes, no manual separation of the subvolumes is needed since the nasal cavity and nasopharyngeal structures are not considered for segmentation.

2.3.5 | Network architecture

CNNs have been widely used in image segmentation and object detection.[29–31] The receptive field and sharing weight characteristics allow CNNs to learn distinctive features from images. In this paper, we use the 2D U-Net as the basic architecture to construct the DL network for all three modes of MRI acquisition utilized in our application. As shown in Figure 8, the network can be divided into two parts: the feature learning modules (left side of U) for extracting multiscale features from the input images, and the feature fusion modules (right side of U) for restoring the spatial information of features and integrating the multiscale features in a feature pyramid. The feature learning modules mainly consist of 10 convolutional layers with 3×3 kernels and 4 max-pooling layers with stride 2. We also incorporate five batch-normalization layers[32] into the architecture for improving the stability and performance of the network. The feature fusion modules are made up of eight convolutional layers, four concatenation layers, and four deconvolutional layers with kernels of size 3×3. To reduce the risk of model overfitting, we added five dropout layers with rate 0.6 into the network.

Due to the small interslice spacing, the adjacent slices in acquired static 3D and isotropic dynamic 3D MRIs have strong spatial correlation. In addition, the shape of the upper airway in mid-retropalatal, mid-retroglossal, and mid-sagittal 2D dynamic MRI varies smoothly over time. To take advantage of the contextual information among adjacent slices in MRI, we utilize spatially/temporally adjacent slices $[I_{t-1}, I_t, I_{t+1}]$ and the annotation data of t-th slice G_t as the input data and GT for training the U-Net network to predict the segmentation result of I_t. The adjacent-slice input strategy enables U-Net to learn the contextual feature from the slices immediately inferior and superior to the target slice, improving the segmentation accuracy. For 2D dynamic MRIs, temporally adjacent slices are utilized. For static 3D and isotropic dynamic 3D MRIs, the three slices denote the three spatially adjacent slices.

2.3.6 | Loss function based on FPs and FNs

Cross-entropy and Dice coefficient are two common loss functions utilized in image segmentation networks. These functions can force the pixel classifier to increase the true positive and true negative rates of the segmentation results during the training stage. However, they are less sensitive specifically to FP and FN classifications. In this paper, to directly handle the segmentation trade-off between FP and FN, we formulate loss as a direct function of FP and FN rates.

Let $S(x)$ and $G(x)$ denote the segmentation result and GT as images, respectively, for an input image $I(x)$. $S(x)$ denotes probabilistic (or fuzzy) value output at pixel x for its classification, and $G(x)$ denotes binary value at pixel x. Then, the loss function to be minimized is defined as:

$$L_{FP,FN}(G, S; W) = \left[\frac{\sum_{x\epsilon\Omega}(1 - G(x))\, S(x; W)}{\sum_{x\epsilon\Omega} S(x; W) + \varepsilon}\right]^2 + \left[\frac{\sum_{x\epsilon\Omega}(1 - S(x; W))\, G(x)}{\sum_{x\epsilon\Omega} G(x) + \varepsilon}\right]^2, \quad (2)$$

where the first and second terms are estimates of FP and FN rates, ε is a constant (set to 0.00001), and W and Ω represent the parameters of the network and image domain, respectively. The derivative of the loss function with respect to W is available, which can be utilized in the back-propagation process. We combined the Dice coefficient loss function with the FP-FN loss function for improving the true positive and true negative rates of

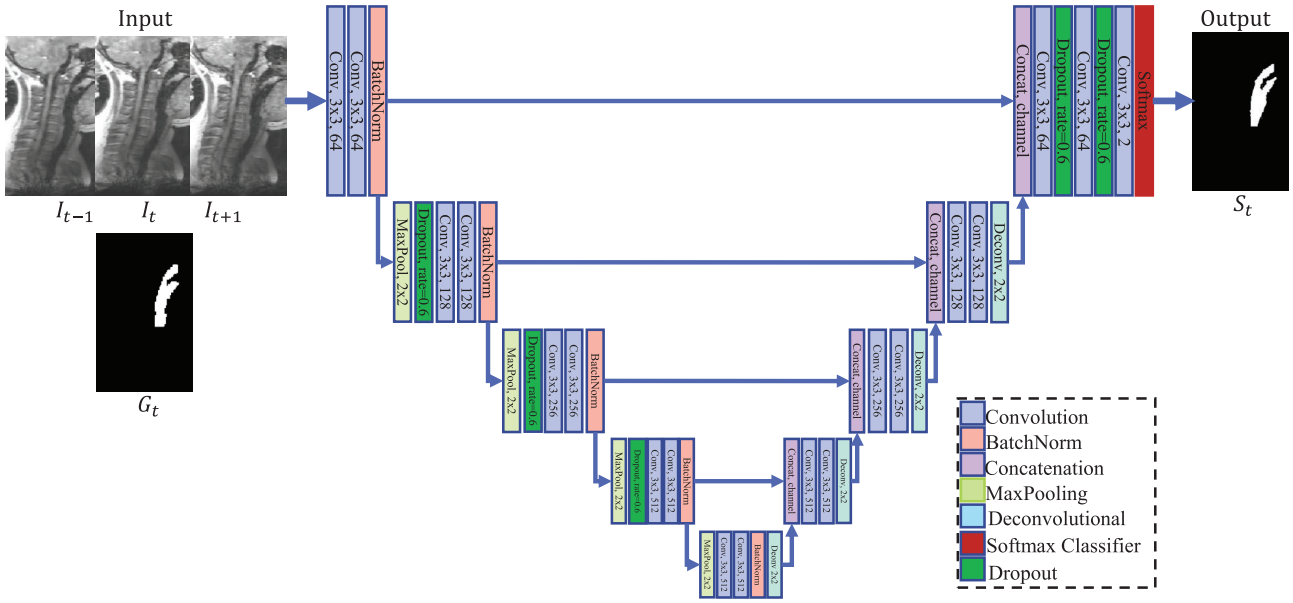

FIGURE 8 The architecture of the upper airway DL network

the segmentation result simultaneously. In addition, we added a weighted regularization term into the loss function to reduce the risk of overfitting. The final loss function employed in our network is:

$$L\ (G, S; W) = L_{FP,FN}\ (G, S; W) - \lambda_1 \times \frac{2\sum_{x\epsilon\Omega} G(x)\, S(x; W)}{\sum_{x\epsilon\Omega} S(x; W) + \sum_{x\epsilon\Omega} G(x) + \varepsilon} + \lambda_2 ||W_1||, \quad (3)$$

where $||\cdot||_1$ is the L1-norm, and λ_1 and λ_2 serve as trade-off parameters among the three terms.

2.3.7 | Data augmentation

It is known that training a DL network requires abundant data for reducing the risk of creating an overfit model. To increase the number and diversity of training data, we employed two data augmentation methods to process each slice in MRI: (1) resizing the image with scale factors 0.8 and 1.2 by a bilinear interpolation method, and (2) varying the intensity level of the image by multiplying the pixel-wise intensity value with scale factors 0.9 and 1.1. Through the data augmentation process, the training data were enlarged to nine times their original size.

3 | EXPERIMENTS, EVALUATION, RESULTS, AND DISCUSSION

3.1 | Experiments

We conducted the following six experiments:

(i) E1: Upper airway segmentation in static 3D MRI. The original MRI was manually separated into two portions: the nasal cavity and nasopharyngeal airway object and the oropharyngeal, hypopharyngeal, and supraglottic/glottic airway object.
(ii) E2: Fully automatic upper airway segmentation in mid-retropalatal and mid-retroglossal dynamic 2D MRI.
(iii) E3: Similar to E2 but in mid-sagittal dynamic 2D MRI.
(iv) E4: Similar to E2 but in isotropic dynamic 3D MRI. In this experiment, the 22 080 3D MRI volumes coming from 23 subjects in the data set were segmented by our approach, including the labeled testing data.
(v) E5: Repeating the above experiments without the GROI strategy.
(vi) E6: Repeating experiment E4 by utilizing cross-entropy and Dice coefficient loss functions to train the DL models.

The above experiments include two stages: (1) training stage for learning and optimizing the parameters of the proposed networks, and (2) inference stage for employing the trained models to process the testing data.

At the training stage, we utilized the Python language and open-source library TensorFlow[33] to implement the DL models. We assigned the trade-off parameters λ_1 and λ_2 in the loss function with values of 0.8 and 10, respectively. We used the Adam optimization algorithm to minimize the loss function with the following hyperparameters: batch size of training data (20), learning rate (0.000001), number of iterations (1000), and dropout rate (0.6). In addition, we employed the trained networks

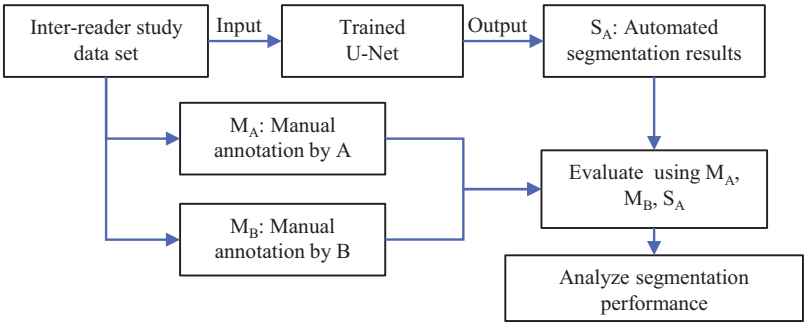

FIGURE 9 A schematic depiction of the inter-reader study

to compute the loss value on the validation data every 100 iterations for searching the optimized parameters for the proposed networks.

At the inference stage, we utilized the GROI strategy and min–max normalization method to process the testing data, inputted the adjacent slices into the trained network to obtain the pixel-wise probability map, and applied a threshold value of 0.5 on the probability map to generate the binary segmentation result. We integrated the five networks together into a single software system for the current OSA application to be able to handle any of the image modalities that are commonly used.

3.2 | Evaluation

To evaluate the segmentation performance of our system, we applied two widely employed metrics on the testing set, including Dice coefficient (DC) and modified Hausdorff distance (HD), defined as follows:

$$\mathrm{DC}(G, S) = \frac{2|S \cap G|}{|S| + |G|}, \tag{4}$$

$$\mathrm{HD}(G, S) = \max\left\{\mathrm{mean}_{g\in G}(inf_{s\in S} d(g, s)), \mathrm{mean}_{s\in S}(inf_{g\in G} d(s, g))\right\}, \tag{5}$$

where $g \in G$ and $s \in S$ denote the pixels/voxels in binary GT and segmentation results as sets, respectively, $d(g, s)$ indicates the distance between g and s, and mean and max indicate the average value and maximum value, respectively. In the classical formulation of HD, the maximum of minimum distances between the segmentation result and GT is estimated. However, such a definition is very sensitive to small FP segmentation regions that are far away from the object. This drawback reduces the reliability of HD. To address this issue, we modified HD by replacing the maximum of minimum by the average of minimum distances.

In addition, to assess the segmentation performance of our system considering the variability in GT, we designed an inter-reader study, as depicted in Figure 9. The data set for this study was annotated by two experts A and B. We denote the training data annotated by A as M_A and by B as M_B. We employ the U-Net networks trained using GT M_A to achieve the segmentation results S_A. Then, we estimate DC(M_A, M_B), DC(M_A, S_A), and DC(M_B, S_A) and compare DC(M_A, M_B) with DC(M_A, S_A), DC(M_A, M_B) with DC(M_B, S_A), and DC(M_A, S_A) with DC(M_B, S_A).

3.3 | Illustrative results

In Figure 10, we show one representative segmentation example of the upper airway from static 3D, mid-retropalatal dynamic 2D, mid-retroglossal dynamic 2D, mid-sagittal dynamic 2D, and isotropic dynamic 3D MRIs. Most of the upper airway segmentation results (shown in orange) have high similarity with the GT (shown in green), especially for the dynamic 2D and dynamic 3D MRIs. By comparing the first and seconds rows in Figure 10, we observe that the segmentation performance for the nasal cavity and nasopharyngeal airway is worse than for the oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway on static 3D MRI. This is because the nasal cavity contains several small and slender air-filled regions that increase the segmentation difficulty for the DL models. Therefore, the segmentation performance for the whole upper airway is better than that for the nasal cavity and nasopharyngeal airway and worse than that for the oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway. By comparing the third with the last column, we note that the FPs and FNs for the DL model with GROI are less than those for the DL model without GROI. In Figure 11, we show two upper airway segmentation examples from isotropic dynamic 3D MRI by the same DL model using different loss functions. We observe that the FPs from the DL model using FP&FN+DICE loss function are less than those using other loss functions.

To demonstrate the 3D segmentation effect of our approach, we utilize CAVASS to display the manual and automatic segmentation results of the static 3D and isotropic dynamic 3D MRIs as 3D surface renderings. Figure 12 shows four representative 3D visualization

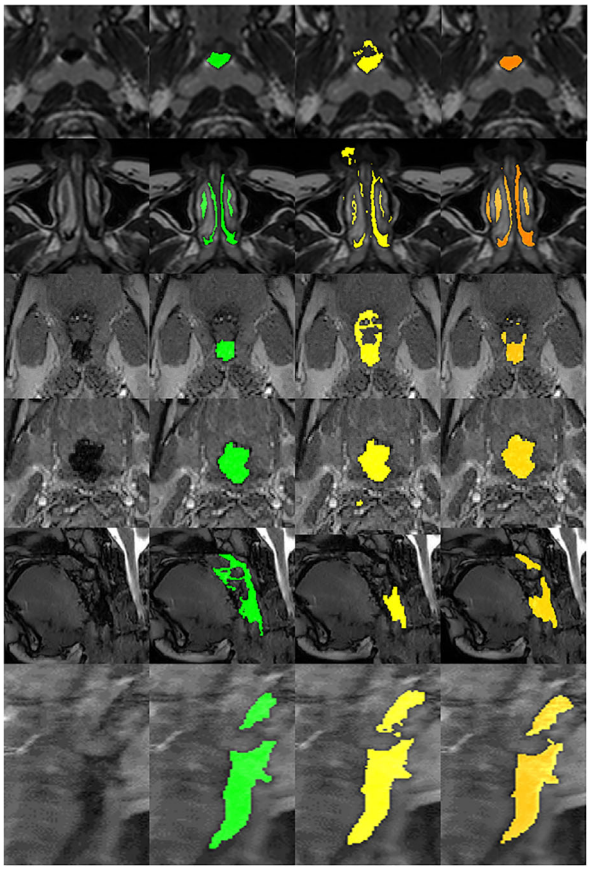

FIGURE 10 Examples of upper airway segmentation. First row: upper airway excluding nasal cavity and nasopharyngeal airway in static 3D MRI. Second row: nasal cavity and nasopharyngeal airway in static 3D MRI. Third row: upper airway in mid-retropalatal dynamic 2D MRI. Fourth row: upper airway in mid-retroglossal dynamic 2D MRI. Fifth row: upper airway in mid-sagittal dynamic 2D MRI. Sixth row: upper airway in isotropic dynamic 3D MRI. The green, yellow, and orange colors represent segmentations from manual, without-GROI, and with-GROI methods, respectively

examples. We observe that the 3D segmentation results show good agreement with the GT, preserving the complete anatomical structure of the upper airway with the details. Considering that each subject dynamic 3D MRI data set contains 480 3D frames in the sleep and wake states, we utilized MATLAB software to batch implement 3D surface rendering for all frames and transfer the frames into video for observing the dynamic variation of upper airway volume. We randomly selected 10 continuous time frames from one subject and displayed in Figure 13. The complete video can be accessed in the Supplementary Data. We observe that the dynamic segmentation results depict the upper airway with its continuously and smoothly varying shape and form. The above results demonstrate the validity and robustness of our segmentation system.

3.4 | Quantitative results

To evaluate the reliability of the GROI strategy, we treated the found ROIs as binary masks and estimated the false-negative rate (FNR) between the GT ROIs and the found ROIs, as shown in Table 4. The FNR on the testing set for all three modes of MR image acquisition was less than 0.001, indicating that the region of the annotated upper airway outside of the found ROIs is exceedingly small. Note that we applied the GROI strategy in the training, validation, and testing stages.

In Table 5, we summarize DC and HD values over the test data sets. The proposed approach achieves acceptably high DC values and low HD values for all MRI modes. It is known that for small and highly sparse objects (like nasal cavity), a DC value of 0.8–0.85 indicates excellent segmentation, while for large space-filling objects, excellent accuracy requires a DC value

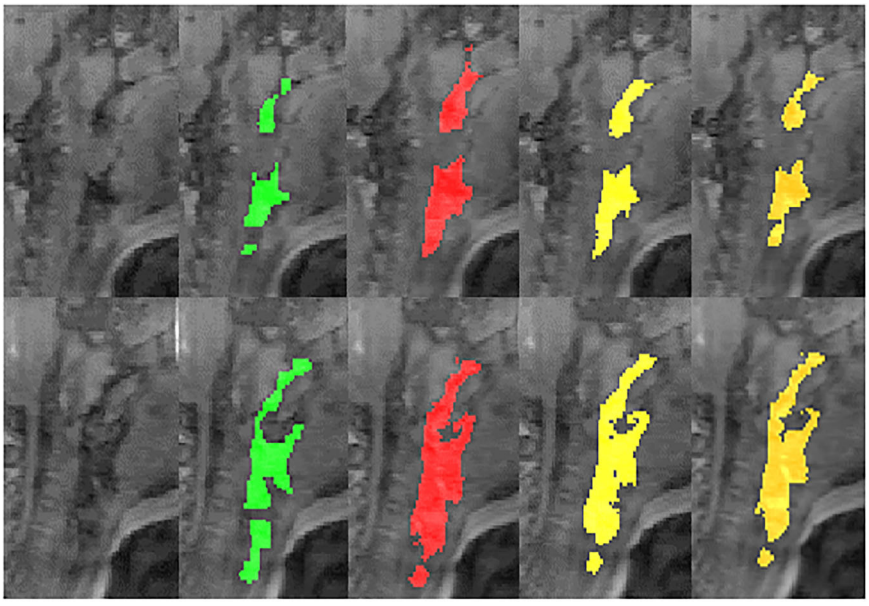

FIGURE 11 Comparison of the impact of different loss functions. First row and second row: upper airway in two slices selected from dynamic 3D MRI from two different subjects. The green, red, yellow, and orange colors represent manual segmentation and results from the DL models trained by cross-entropy, Dice coefficient, and FP&FN+DICE loss functions, respectively

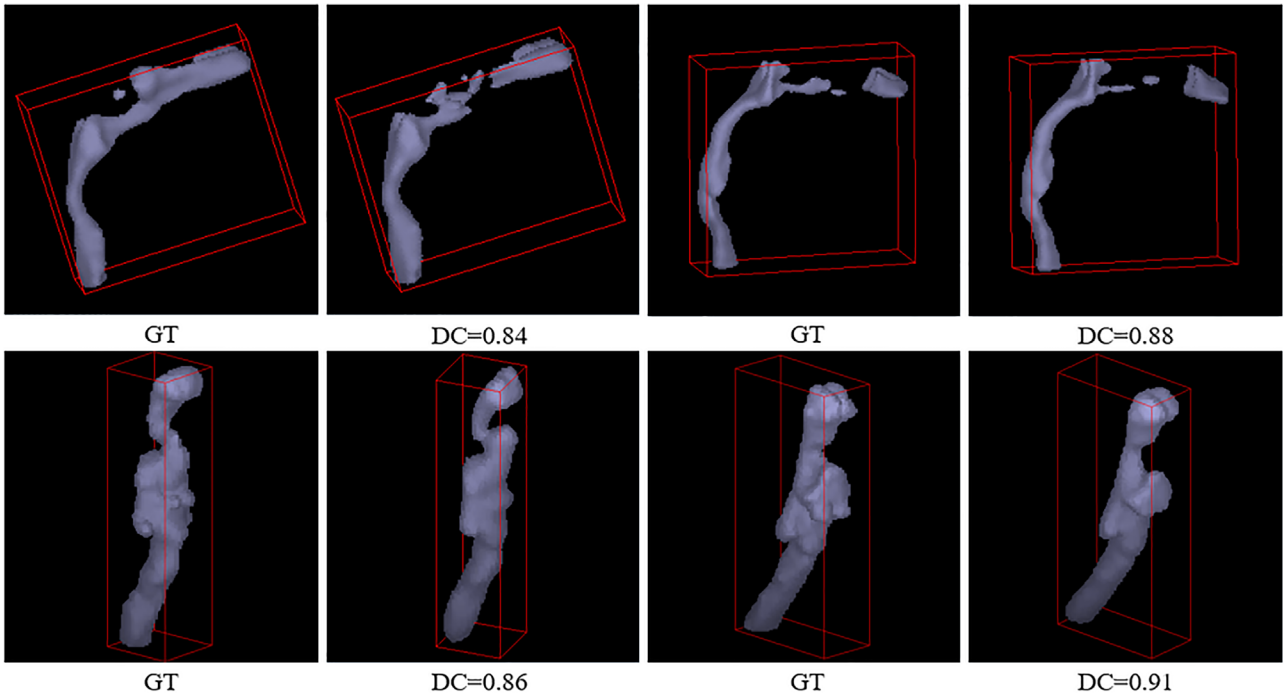

FIGURE 12 Illustration of upper airway 3D segmentation result: first row: from static 3D MRI. Second row: from isotropic dynamic 3D MRI. First and third columns: GT segmentation. Second and fourth columns: automatic segmentation result. These results were from four subjects

FIGURE 13 Illustration of 10 continuous time frames in a sample upper airway 4D segmentation result

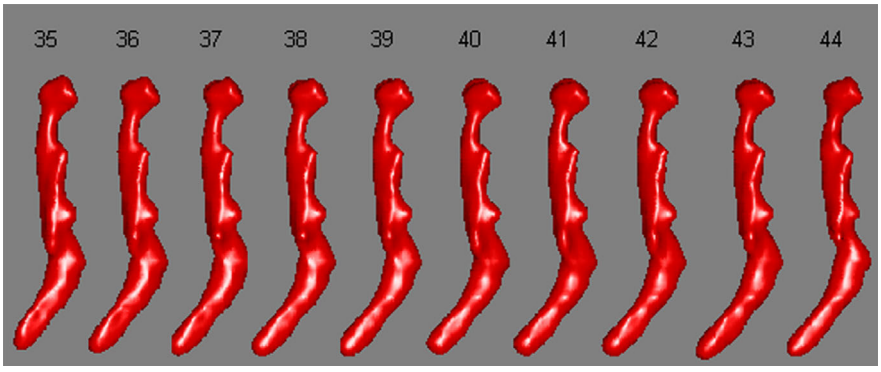

TABLE 4 Parameters of GROI for the different MRI protocols

	GROI parameters		
Data type (mode)	**center point**	**window size**	**FNR on testing set**
Static 3D MRI—nasal cavity and nasopharyngeal airway	(0.51, 0.25)	96 × 256	< 0.001
Static 3D MRI—oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway	(0.49, 0.44)	96 × 128	< 0.001
Mid-retropalatal and mid-retroglossal dynamic 2D MRI	(0.51, 0.48)	96 × 96	< 0.001
Mid-sagittal dynamic 2D MRI	(0.60, 0.62)	112 × 128	< 0.001
Isotropic dynamic 3D MRI	(0.67, 0.32)	96 × 192	< 0.001

TABLE 5 Segmentation accuracy on test data expressed as mean/standard deviation

		Without GROI		With GROI		
Data type	**Object**	**DC**	**HD (pixels)**	**DC**	**HD (pixels)**	***p*-Value between DC with and without GROI. Second *p*-value is for HD**
Static 3D MRI	Nasal cavity and nasopharyngeal airway	0.75/0.03	3.26/3.13	0.80/0.03	2.11/0.86	< 0.001
						0.004
	Oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway	0.87/0.04	2.98/2.71	0.90/0.03	2.19/1.85	< 0.001
						0.005
	Whole upper airway	0.81/0.03	2.33/2.25	0.84/0.03	1.99/0.62	< 0.001
						0.08
Mid-retropalatal and mid-retroglossal dynamic 2D MRI	Upper airway	0.84/0.14	3.44/6.26	0.89/0.13	2.10/2.15	< 0.001
						< 0.001
Mid-sagittal dynamic 2D MRI	Upper airway	0.79/0.11	6.53/10.55	0.84/0.07	3.72/1.86	< 0.001
						< 0.001
Isotropic dynamic 3D MRI	Upper airway	0.79/0.07	4.48/5.26	0.86/0.05	3.18/2.91	< 0.001
						< 0.001

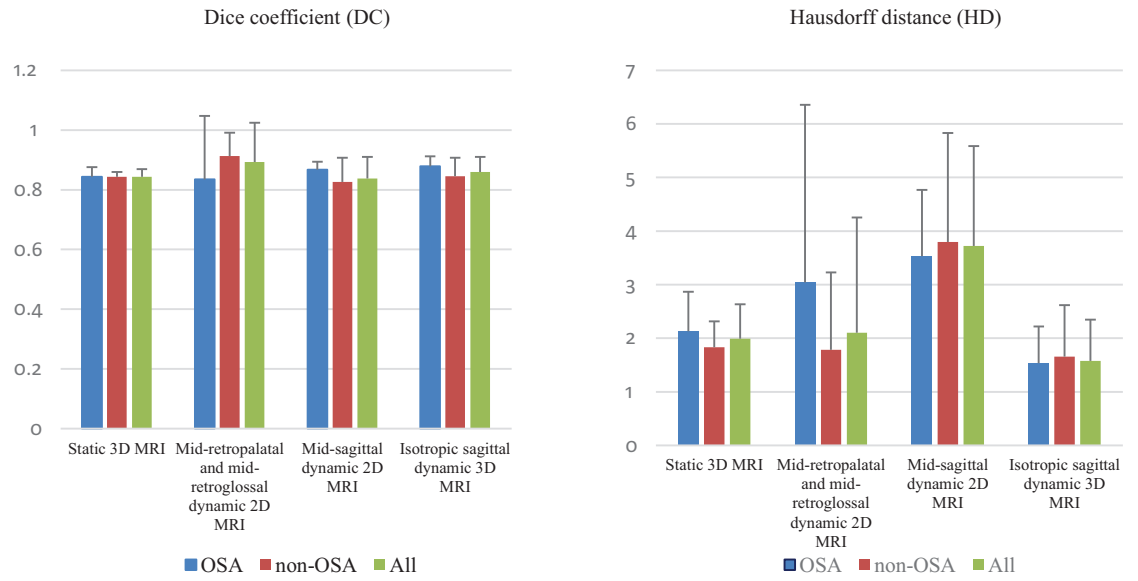

FIGURE 14 DC and HD (in pixels) for upper airway in OSA, non-OSA, and all subjects. The bars represent mean values, and the whiskers denote standard deviations

of > 0.9.[34] In addition, the low standard deviation in DC indicates that the proposed approach has good stability and robustness. To explore the difference in performance in OSA and non-OSA subjects, we compare in Figure 14 results from these groups. From these and other results not shown here, we find that our approach achieves good accuracy for both groups and the difference between DC in OSA and non-OSA subjects is not statistically significant.

Comparison between DC values obtained with and without the use of GROI indicates that the proposed anatomy guidance via GROI improves segmentation. The sharp decline of HD values means that some FP pixels far away from the GT are removed by the GROI strategy. The improvements in DC and HD are statistically significant. In Table 6, we show results from the use of different loss functions. The proposed FP&FN with DICE loss function achieves the highest DC value

TABLE 6 Segmentation accuracy (mean/standard deviation) by using different loss functions

Data type	Loss function type	Evaluation metrics				*p*-Value for comparing DC	
		TP	FP	DC	HD	Cross-entropy versus FP&FN + DICE	DICE versus FP&FN + DICE
Isotropic dynamic 3D MRI	Cross-entropy	0.92/0.05	0.0027/0.0011	0.81/0.04	2.22/0.57		
	DICE	0.81/0.09	0.0008/0.0007	0.84/0.05	3.68/2.46		
	FP&FN+DICE	0.83/0.06	0.0007/0.0005	0.86/0.05	3.18/2.91	< 0.001	< 0.001

TABLE 7 Inter-reader study results showing three DC values expressed as mean/standard deviation and the *p* value for the three pairwise comparisons

Object	DC(M_A, M_B)	DC(M_A, S_A)	DC(M_B, S_A)	*p* Value DC(M_A, M_B) versus DC(M_A, S_A)	*p* Value DC(M_A, M_B) versus DC(M_B, S_A)	*p* Value DC(M_A, S_A) versus DC(M_B, S_A)
Nasal cavity and nasopharyngeal airway in static 3D MRI	0.83/0.06	0.82/0.06	0.77/0.06	0.85	0.06	0.07
Oropharyngeal, hypopharyngeal, and supraglottic/glottic laryngeal airway in static 3D MRI	0.88/0.04	0.89/0.10	0.85/0.09	0.69	0.39	0.34
Upper airway in mid-retropalatal and mid-retroglossal dynamic 2D MRI	0.80/0.13	0.85/0.17	0.80/0.16	< 0.001	0.74	< 0.001
Upper airway in mid-sagittal dynamic 2D MRI	0.77/0.09	0.87/0.07	0.76/0.09	< 0.001	0.024	< 0.001
Upper airway in isotropic dynamic 3D MRI	0.57/0.09	0.85/0.03	0.53/0.10	< 0.001	0.25	< 0.001

and lowest FP rate, meaning that the FP&FN-based loss term is helpful to reduce the FPs. The difference in segmentation performance among the three loss functions is statistically significant ($p < 0.001$).

It is known that DC behaves nonlinearly with respect to sparse and nonsparse objects;[34] that is, a lower DC for sparse and small objects indicates the same level of segmentation quality as a higher DC for large nonsparse objects. Additionally, intra and interoperator variability in GT is known to be much higher for sparse objects (with a DC comparing different GT segmentations in the range 0.8–0.85) than for nonsparse objects (DC > 0.9).[34] We postulated that our current results are as good as GT, meaning that they are statistically indistinguishable from GT if variability in GT is taken into consideration.

To verify this hypothesis, we estimated GT variability in our application by conducting the inter-reader study outlined above. The results are summarized in Table 7. We note that DC(M_A, S_A) is slightly lower than DC(M_A, M_B) in static MRI, although the difference is statistically not significant ($p = 0.85, 0.69$), demonstrating that our autosegmentation in static 3D MRI is as good as GT. Not surprisingly, the other two comparisons approach statistical significance ($p = 0.06$ and 0.07) for the nasal cavity and nasopharyngeal airway, suggesting that the variation between A and B is large enough for these subtle objects to show significantly better performance with A than with B.

For all dynamic modes, our system trained on M_A performs statistically significantly better ($p < 0.001$) than the variability observed between A and B. The results are also statistically significantly better ($p < 0.001$) when M_B is used as reference to evaluate accuracy, suggesting that the outputs of our system favor A. In summary, these upper airway structures are very subtle, and it is very difficult to consistently create highly reliable GT. As such, given the high variability in GT, it is very difficult to achieve DC > 0.9, and the results will favor the GT(s) that are utilized for training the models.

3.5 | Computational considerations

All experiments were conducted on a PC with an Intel i7-7700K CPU and two NVIDIA 1080 Ti GPUs. We aimed to construct a unified system with high accuracy and efficiency for segmenting the upper airway in static and dynamic 2D and 3D MRI scans. To assess the efficiency of our system, we compute the average segmentation time per subject and slice, as shown in Table 8. We observe that the segmentation time per subject and per slice of our approach for static 3D and dynamic 2D MRI is lower than 1.2 and 0.01 s, respectively. Note that the

TABLE 8 Segmentation time in seconds (per subject and slice) for our approach

Data type	Per subject	Per slice
Static 3D MRI	0.20	0.006
Mid-retropalatal and mid-retroglossal dynamic 2D MRI	1.19	0.006
Mid-sagittal dynamic 2D MRI	1.09	0.009
Isotropic dynamic 3D MRI	396.93	0.009

average segmentation time for static 3D MRI does not include the time for manual parcellation of upper airway, which typically is about 10 s. The 960 3D volumes of one subject in the isotropic dynamic 3D MRI data set can be segmented in less than 6.7 min. The efficiency of our approach even in this extreme case of high data volume can meet the practical needs of image analysis in OSA for clinical or research purposes.

3.6 | Comparison with published literature

We compare our approach with several state-of-the-art methods as summarized in Table 9. Our work has several advantages over published studies:

(i) Data type: The proposed approach can be employed with three modes of MR image acquisition with maximum demonstrable accuracy and robustness. In contrast, other methods are applicable to only one specific MRI protocol. Considering the significant discrepancies among the different modes of MR image acquisition, we designed a general method and constructed five networks with the same architecture. This strategy allowed each DL model to focus on the specific MRI protocol, improving the segmentation performance for our approach.

(ii) Data size: In our work, we collected and annotated data sets from 160 subjects with 20 048 MRI slices as the experimental data, which is much greater than that of other studies. In particular, the isotropic dynamic 3D MRI data set includes 23 subjects with 229 annotated MRI frames (3D volumes). We randomly selected 10 subjects, including 164 annotated frames, and 10 subjects, including 40 annotated frames from the subdata set for training and testing our approach. In comparison, the largest data set in the published literature[23] contains 36 training and 9 testing cases. It is known that less training data result in the strong possibility of model overfitting and decrease the generalizability and confidence of model evaluation.

(iii) Anatomy definition: Distinct from other studies, we provide the precise anatomical definition of the upper airway body region in three modes of MR image acquisition, enabling uniformity of the approach in all applications. In addition, we trained two networks to segment two subobjects of the upper airway in static MRI according to anatomical parcellation of the upper airway. These design principles effectively improve segmentation accuracy.

(iv) OSA application: The acquired static and dynamic MRI data set includes 56 subjects with OSA. The

TABLE 9 Comparison with methods in the literature

Approach	Data type (image contrast)	Number of train/test cases	Object	DC (mean/standard deviation or minimum–maximum)
Ivanovska et al.[17]	Static MRI (T1 weighted)	*NA*/6	Pharynx	0.86/3.13
Javed et al.[18]	Sagittal real-time MRI	*NA*/10	Upper airway	0.84–0.94
Shahid et al.[19]	Static MRI (T1 weighted)	*NA*/16	Pharynx	0.89/*NA*
Silva and Teixeira[20]	Sagittal real-time MRI (T1 weighted)	51/50 slices	Vocal tract	0.82/0.05
Li et al.[21]	Sagittal cine MRI	10/3	Upper airway	0.93/1.46
Erattakulangara and Lingala[22]	Sagittal 2D MRI	75/20 slices	Vocal tract	0.90/*NA*
Liu et al.[23]	CT and synthetic MRI (T1 weighted)	36/9	Pharynx	0.74/0.10
Proposed method	Static 3D MRI (T2 weighted)	30/22	Upper airway	0.84/0.03
	Mid-retropalatal and mid-retroglossal dynamic 2D MRI (balanced SSFP, T1/T2 weighted)	2073/1056 slices	Upper airway	0.89/0.13
	Mid-sagittal dynamic 2D MRI (balanced SSFP, T1/T2 weighted)	1756/1041 slices	Upper airway	0.84/0.07
	Isotropic dynamic 3D MRI (gradient echo, T2* weighted)	164/40 frames	Upper airway	0.86/0.05

Abbreviation: NA, information not available.

proposed approach achieves good segmentation performance for both normal pediatric subjects and pediatric patients with OSA. In contrast, other studies did not demonstrate the reliability of segmentation for subjects with OSA. Therefore, our approach is more applicable in OSA investigation.

(v) Segmentation performance: Due to the differences between test sets used in our approach and other studies from the literature, a direct comparison of performance is unavailable from Table 8. Main factors potentially influencing the segmentation performance include the number of test cases, the quality of test cases, the definition of the upper airway, the quality and variability of manual annotation, and so on. We utilized the largest and independent data set to estimate the metric values of performance for our approach. The mean value of DC based on our approach is competitive with that of other studies, and its standard deviation is lower than that of most other methods.

3.7 | Limitations

There are some limitations in our approach. First, the segmentation approach for the static 3D MRI requires users to provide the initial location of the inferior, separation, and superior boundary slices of the upper airway. We are working on an automatic parcellation method to implement this step. Second, it is difficult for our approach to segment the slender regions of the nasal cavity in static MRI, resulting in disconnected air passages in the result. A pixel-wise weighting strategy in the loss function to increase the importance of the intractable segmented regions is being tested. Third, the parameters of GROI, learned from the training data, may not be applicable to other test MRI data sets due to differences in data acquisition from different centers. A more general framework from body-wide AAR[26,35] should further facilitate automatically localizing the upper airway in MRI regardless of differences in data acquisition based purely on geographic anatomic layout.

4 | CONCLUSIONS

In this paper, we propose a novel comprehensive system for automatic upper airway segmentation in static 3D, dynamic 2D, and dynamic 3D MRI for the study of patients with OSA. Our approach is mainly comprised of two parts: (1) a judicious design of methodology combining human-knowledge and DL techniques to tackle this incredibly challenging segmentation problem. (2) A careful adaptation of the 2D U-Net network with a novel loss function to try to achieve the maximum possible performance for the OSA application given the low quality of the images and low contrast and very subtle/sparse nature of the structures. The experimental results on a large and independent testing data set demonstrate that our approach achieves excellent agreement with manual segmentation and becomes indistinguishable from it, often achieving better robustness than manual GT segmentation. In addition, our approach shows several advantages compared to those in the published literature as explained above. The proposed approach has the potential to be utilized in other dynamic MRI-related applications, such as for lung or heart segmentation from dynamic MRI following similar design principles.

ACKNOWLEDGMENTS

This work is supported by an NIH grant HL130468 and in part by the Institute for Translational Medicine and Therapeutics of the University of Pennsylvania through a grant by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR001878. The visit of Mr. Lipeng Xie to the Medical Image Processing Group, Department of Radiology, University of Pennsylvania, is supported by China Scholarship Council.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Research data are not shared.

REFERENCES

1. Eckert DJ. Phenotypic approaches to obstructive sleep apnoea—new pathways for targeted therapy. *Sleep Med Rev*. 2018;37:45-59.
2. Osman AM, Carter SG, Carberry JC, Eckert DJ. Obstructive sleep apnea: current perspectives. *Nat Sci Sleep*. 2018;10:21.
3. Zinchuk AV, Jeon S, Koo BB, et al. Polysomnographic phenotypes and their cardiovascular implications in obstructive sleep apnoea. *Thorax*. 2018;73(5):472-480.
4. Wang Y, Sun C, Cui X, Guo Y, Wang Q, Liang H. The role of drug-induced sleep endoscopy: predicting and guiding upper airway surgery for adult OSA patients. *Sleep Breath*. 2018;22(4):925-931.
5. Caballero P, Alvarez-Sala R, García-Río F, et al. CT in the evaluation of the upper airway in healthy subjects and in patients with obstructive sleep apnea syndrome. *Chest*. 1998;113(1):111-116.
6. Galvin JR, Rooholamini SA, Stanford W. Obstructive sleep apnea: diagnosis with ultrafast CT. *Radiology*. 1989;171(3):775-778.
7. Fleck RJ, Shott SR, Mahmoud M, Ishman SL, Amin RS, Donnelly LF. Magnetic resonance imaging of obstructive sleep apnea in children. *Pediatr Radiol*. 2018;48(9):1223-1233.
8. Gamaleldin O, Bahgat A, Anwar O, et al. Role of dynamic sleep MRI in obstructive sleep apnea syndrome. *Oral Radiol*. 2020;37:1-9.
9. Slaats MA, Van Hoorenbeeck K, Van Eyck A, et al. Upper airway imaging in pediatric obstructive sleep apnea syndrome. *Sleep Med Rev*. 2015;21:59-71.
10. Wagshul ME, Sin S, Lipton ML, Shifteh K, Arens R. Novel retrospective, respiratory-gating method enables 3D, high resolution, dynamic imaging of the upper airway during tidal breathing. *Magn Reson Med*. 2013;70(6):1580-1590.
11. Liu J, Udupa JK, Odhnera D, McDonough JM, Arens R. System for upper airway segmentation and measurement with MR

imaging and fuzzy connectedness. *Acad Radiol*. 2003;10(1): 13-24.
12. Udupa JK, Samarasekera S. Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. *Graph Models Image Proc*. 1996;58(3):246-261.
13. Tong Y, Udupa JK, Odhner D, et al. Minimally interactive segmentation of 4D dynamic upper airway MR images via fuzzy connectedness. *Med Phys*. 2016;43(5):2323-2333.
14. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
15. Sun C, Udupa JK, Tong Y, et al. Segmentation of 4D images via space-time neural networks. Paper presented at: Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging. 2020.
16. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015.
17. Ivanovska T, Dober J, Laqua R, Hegenscheid K, Völzke H. Pharynx segmentation from MRI data for analysis of sleep related disoders. Paper presented at: International Symposium on Visual Computing. 2013.
18. Javed A, Kim Y, Khoo MCK, Ward SLD, Nayak KS. Dynamic 3-D MR visualization and detection of upper airway obstruction during sleep using region-growing segmentation. *IEEE Trans Biomed Eng*. 2015;63(2):431-437.
19. Shahid MLUR, Chitiboi T, Ivanovska T, et al. Automatic pharynx segmentation from MRI data for obstructive sleep apnea analysis. Paper presented at: VISAPP. 2015.
20. Silva S, Teixeira A. Unsupervised segmentation of the vocal tract from real-time MRI sequences. *Comput Speech Lang*. 2015;33(1):25-46.
21. Li H, Chen HC, Dolly S, et al. An integrated model-driven method for in-treatment upper airway motion tracking using cine MRI in head and neck radiation therapy. *Med Phys*. 2016;43(8Part1):4700-4710.
22. Erattakulangara S, Lingala SG. Airway segmentation in speech MRI using the U-net architecture. Paper presented at: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). 2020.
23. Liu Y, Lei Y, Fu Y, et al. Head and neck multi-organ auto-segmentation on CT images aided by synthetic MRI. *Med Phys*. 2020;47(9):4294-4302.
24. Xie L, Udupa JK, Tong Y, et al. Automatic upper airway segmentation in static and dynamic MRI via deep convolutional neural networks. Paper presented at: Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging. 2021.
25. Grevera G, Udupa J, Odhner D, et al. CAVASS: a computer-assisted visualization and analysis software system. *J Digit Imaging*. 2007;20(1):101.
26. 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.
27. Zhuge Y, Udupa JK. Intensity standardization simplifies brain MR image segmentation. *Comput Vis Image Underst*. 2009;113(10):1095-1103.
28. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging*. 2000;19(2):143-150.
29. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84-90.
30. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42: 60-88.
31. Hu Z, Tang J, Wang Z, Zhang K, Zhang L, Sun Q. Deep learning for image-based cancer detection and diagnosis: a survey. *Pattern Recognit*. 2018;83:134-149.
32. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Paper presented at: International Conference on Machine Learning. 2015.
33. Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning. Paper presented at: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI}. 2016.
34. Li J, Udupa JK, Tong Y, Wang L, Torigian DA. LinSEM: linearizing segmentation evaluation metrics for medical images. *Med Image Anal*. 2020;60:101601.
35. 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-and-neck and thoracic cancer cases. *Med Image Anal*. 2019;54:45-62.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Xie L, Udupa JK, Tong Y, et al. Automatic upper airway segmentation in static and dynamic MRI via anatomy-guided convolutional neural networks. *Med. Phys.* 2022;49:324–342.
https://doi.org/10.1002/mp.15345