AMERICAN ACADEMY OF
OTOLARYNGOLOGY–
HEAD AND NECK SURGERY
FOUNDATION

Otolaryngology–
Head and Neck Surgery
2024, Vol. 00(00) 1–9
© 2024 American Academy of
Otolaryngology–Head and Neck
Surgery Foundation.
DOI: 10.1002/ohn.868
http://otojournal.org
WILEY

A Label-Efficient Framework for Automated Sinonasal CT Segmentation in Image-Guided Surgery

Manish Sahu, PhD[1]*, Yuliang Xiao, MSE[1]*, Jose L. Porras, MD[2], Ameen Amanian, MD, MSE[3], Aseem Jain, BS[1], Andrew Thamboo, MD[3], Russell H. Taylor, PhD[1], Francis X. Creighton, PhD[4], and Masaru Ishii, PhD, MD[4]

Abstract

Objective. Segmentation, the partitioning of patient imaging into multiple, labeled segments, has several potential clinical benefits but when performed manually is tedious and resource intensive. Automated deep learning (DL)-based segmentation methods can streamline the process. The objective of this study was to evaluate a label-efficient DL pipeline that requires only a small number of annotated scans for semantic segmentation of sinonasal structures in CT scans.

Study Design. Retrospective cohort study.

Setting. Academic institution.

Methods. Forty CT scans were used in this study including 16 scans in which the nasal septum (NS), inferior turbinate (IT), maxillary sinus (MS), and optic nerve (ON) were manually annotated using an open-source software. A label-efficient DL framework was used to train jointly on a few manually labeled scans and the remaining unlabeled scans. Quantitative analysis was then performed to obtain the number of annotated scans needed to achieve submillimeter average surface distances (ASDs).

Results. Our findings reveal that merely four labeled scans are necessary to achieve median submillimeter ASDs for large sinonasal structures—NS (0.96 mm), IT (0.74 mm), and MS (0.43 mm), whereas eight scans are required for smaller structures—ON (0.80 mm).

Conclusion. We have evaluated a label-efficient pipeline for segmentation of sinonasal structures. Empirical results demonstrate that automated DL methods can achieve submillimeter accuracy using a small number of labeled CT scans. Our pipeline has the potential to improve preoperative planning workflows, robotic- and image-guidance navigation systems, computer-assisted diagnosis, and the construction of statistical shape models to quantify population variations.

Level of Evidence. N/A

Keywords

automated segmentation and registration, deep learning, image-guided surgery, medical image segmentation, semi-supervised learning, sinonasal computed tomography

Received December 12, 2023; accepted June 3, 2024.

Functional endoscopic sinus surgery (FESS) is an effective, minimally invasive procedure for the treatment of chronic inflammatory conditions of the paranasal sinuses that offers minimal pain and no outward scarring of the nose. Despite FESS being performed over 250,000 times annually in the United States, it remains a challenging procedure owing to narrow operative corridors and the proximity of critical anatomical structures.[1] Although both surgeon experience and advances in imaging have contributed to decreased rates of FESS complications, the reported rates of major complications range between 0.31% and 0.47% while minor complications are reported at rates between 1.37% and 5.6%.[2]

The use of intraoperative navigation in sinonasal surgery registered to a patient's preoperative imaging

[1]Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, Maryland, USA
[2]Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
[3]Division of Otolaryngology, Department of Surgery, University of British Columbia, Vancouver, British Columbia, Canada
[4]Department of Otolaryngology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
*These authors contributed equally to this article.

Presented at the Triologic Society meeting at Combined Otolaryngology Spring Meetings (COSM); May 4-6, 2023; Boston, Massachusetts

Corresponding Author:
Jose L. Porras, MD, Department of Neurosurgery, Johns Hopkins University School of Medicine, Hospital 6007, 1800 Orleans St Sheikh Zayed Tower, Baltimore, MD 21287, USA.
Email: jporras1@jhmi.edu

can decrease operative times and improve outcomes.[3] Intraoperative navigation is especially helpful in patients with distorted anatomy or a burden of disease that cannot be appreciated on endoscopic visual inspection alone. However, current navigation systems suffer from tracking errors greater than 1 mm, which, when considering the relative size of sinonasal structures, is a significant limitation.[4-6] Errors of this magnitude also limit the use of preoperative labeling of anatomical structures on patient imaging, a process referred to as segmentation. With a robust system for segmentation, the surgical team would be assisted in augmenting the preoperative planning workflow or in developing patient-specific anatomical models for image- or robot-guided navigation systems.

Currently, manual segmentation of anatomical structures in volumetric imaging is often required for such systems, meaning that radiologists or other qualified personnel must invest hours of time in labeling structures. To reduce the burden of manual labeling, semi- or fully automated segmentation methods have been proposed.[7] Semi-automated labeling has been a feature of atlas-based labeling methods,[8] the traditional technique used for segmentation tasks.[7,9] Atlas-based methods rely on creating an atlas or a reference volume, which is generated from the segmentation of a single image or the average of multiple images. This presegmented atlas is then coregistered with a patient CT to automatically segment the patient's anatomy. For instance, Konuthula et al[9] proposed a semi-automated atlas-based method for segmenting skull base structures. The technique utilizes rigid landmark registration followed by a deformable registration algorithm[10] to achieve semi-automated segmentation. Major limitations of such atlas-based segmentation methods include extended execution times, challenges in accommodating patient anatomy variation, and the requirement of large, annotated datasets to generate accurate atlases.[7,11] Recently, deep learning (DL) based methods, such as U-Net,[12] have demonstrated impressive performance on various medical image segmentation tasks.[13,14] In contrast to atlas-based approaches, DL-based auto-segmentation requires considerably less time for on-line applications.[7,8] However, DL performance relies on large, annotated scan databases. DL method performance declines significantly when ground-truth segments are limited, which is often the case in a healthcare context since manual annotation is a severely time-consuming process for radiologists.

To overcome these obstacles, our group has developed label-efficient segmentation models to decrease dependence on manual labeling for establishing ground-truths. The present study proposes an annotation-efficient segmentation pipeline that involves training a DL method with a small subset of manually labeled data along with a pool of unlabeled data (a technique referred to as semi-supervised learning) to generate accurate and efficient segmentation masks of the critical sinonasal structures on CT. Motivated by the 1 mm navigation error tolerance for intraoperative image guidance in sinus surgery,[5,6] we analyzed the number of labeled scans needed to reach submillimeter labeling accuracy.

Materials and Methods

Ethics

This study was approved by the Johns Hopkins School of Medicine Institutional Review Board. To prepare the dataset for deep neural network training, we obtained high-resolution CT scans from the database of a tertiary referral Otolaryngology center for adult patients (>18 years of age). Prior to manual annotation of anatomical structures, all CT images were deidentified by excluding the soft tissue structures of the face.

Dataset Creation and Manual Segmentation

The resolution and dimension of each scan is 0.46 mm per voxel length and 512 × 512 × N respectively, where N represents the number of slices in the axial direction of CT image. Our project sought to identify nonpathological sinonasal anatomic structures and therefore, patients with identified sinonasal pathologies (eg, polyp, tumor, or prior trauma) were not included in the dataset. In total, 40 scans satisfied our inclusion criteria, of which 16 scans were labeled manually via an open-source medical imaging software, *3D Slicer*.[15] The anatomical structures (eg, nasal septum [NS], inferior turbinate [IT], maxillary sinus [MS], and optic nerve [ON]) (**Figure 1**) were annotated by both a senior Otolaryngology–Head and Neck Surgery resident and a medical trainee with prior knowledge of sinonasal anatomy. Final annotations were verified by the senior author (rhinology and skull base surgeon). It is worth noting that the dataset for model training was comprised of a small number of manually annotated CT scans combined with unlabeled CT scans, while the remaining labeled CT scans served as a validation set for the evaluation process.

Dataset Processing

To facilitate effective joint training of the DL model with both labeled and unlabeled scans, the anonymized scans were rigidly co-registered using an open-source deformable image registration software, Advanced Normalization Tools (ANTs[16]), to standardize geometry, dimensions, and coordinate spaces. Prior to training, a region of interest (224 × 224 × 224) was cropped from each scan based on the template scan range used for ANTs.

DL Framework

Our framework is built on the DL method, *DeepAtlas*[17] which enables label-efficient learning by jointly training two DL models—the segmentation model and the registration model. There are two key features of this method: its ability to learn from only a few manual segmentations and its use

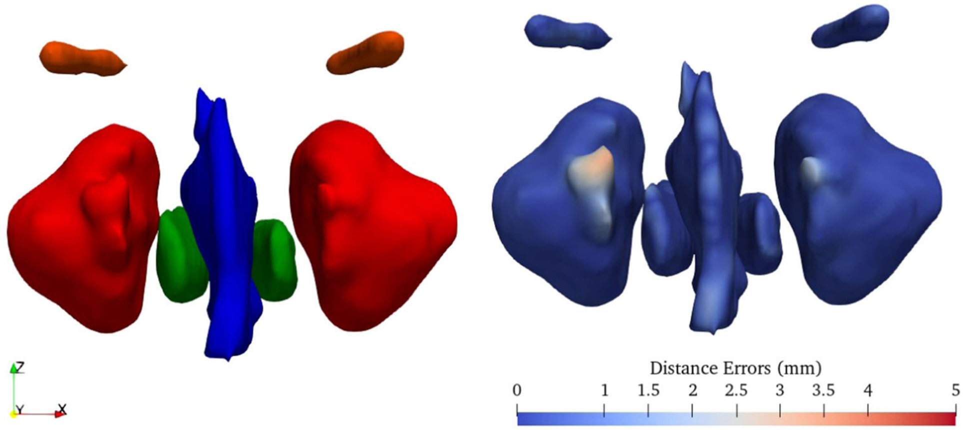

Figure 1. Visual comparison between a sample ground-truth segmentation (left) and heatmap visualization of predicted segmentations (right) across datasets. The range of distance errors lies between 0 mm (blue) and 5 mm (red).

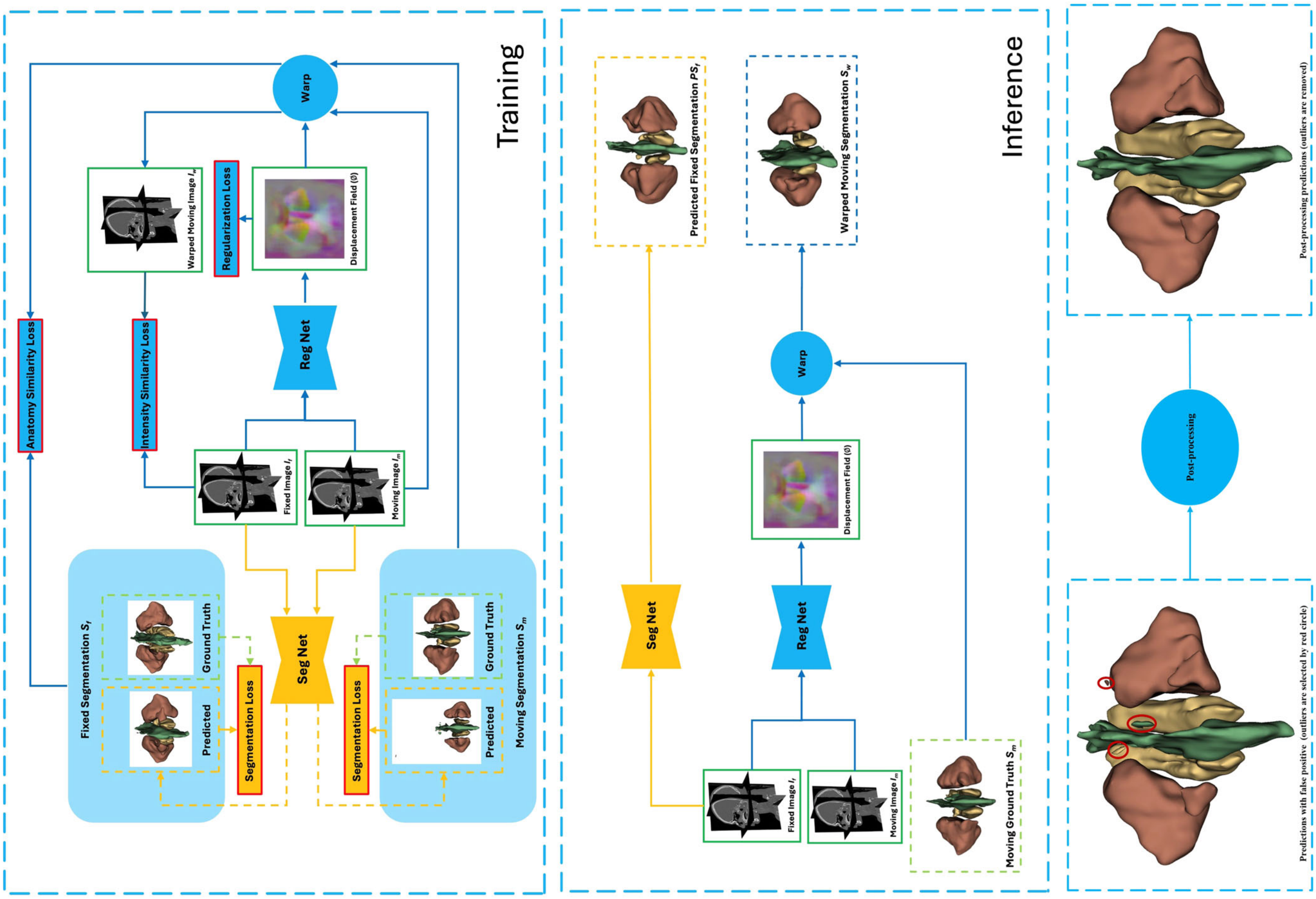

Figure 2. Overview of our automated segmentation pipeline including network training (top), network inference (middle), and postprocessing (bottom). In the training phase (top), the two DL models iteratively learn from a pair of randomly selected scans. During inference for the registration network (middle), the labels associated with a moving image are spatially transformed to the target image using the learned transformation. This technique is commonly known as label propagation from the moving image to the target image. (bottom) To remove false positive generated by our deep learning models, we employ a post-processing step to identify and subsequently remove the small island regions.

of unlabeled data for mutual guidance of the registration or segmentation model, respectively.

Prior to model training, users can specify the number of segmentations to use. The training process follows an iterative approach (see **Figure 2**) where the DL models learn from a pair of images, denoted as fixed and moving images, along with their segmentation labels when available. The registration model takes these image pairs and generates a deformation field, which spatially transforms each voxel of the moving image to align with the fixed image. Subsequently, this deformation field is applied to both the moving image and its corresponding segmentation to produce deformed scans and segmentations for the fixed image. To facilitate effective training, we employ two types of losses: one measures the similarity between the fixed and deformed scans, while the other assesses the consistency between the estimated and ground truth segmentations. Additionally, a regularization loss, dependent on the deformation field, promotes smooth network training. Concurrently, the individual paired images are fed into the segmentation model, which estimates a segmentation map for each image. If sufficient ground truth data is available, we compute a segmentation similarity loss for each input scan; otherwise, this step is omitted. Subsequently, the losses for both networks are combined and minimized iteratively until no further decrease in loss is observed.

Overall, this iterative approach enables the segmentation model to generate a voxel-based probability map from a few labeled scans, while the registration model learns a deformation field needed to register a pair of image scans. Thus, the joint approach combines the strengths of both the atlas-based registration method and DL-based segmentation model in a unified manner, where both models mutually guide each other's training in a label-efficient manner. Both the registration and the segmentation model were implemented using a 3-dimensional (3D) *U-Net* architecture.[18] At the end of the training, the registration and segmentation models can independently make predictions on test scans.

Postprocessing

The predictions of the trained DL models often generate false positives in the form of small islands around expected anatomical structures. To remove these false positives, we use the 3D connected component algorithm (CC3D).[19] Connected components groups segmented voxels of anatomical structures into spatial regions based on voxel connectivity. After identifying these spatial regions, we then refine the algorithm's segmentation by automatically removing smaller regions (**Figure 2**).

Segmentation Evaluation Metrics

To evaluate performance of our DL framework, the predictions from the registration and segmentation models were evaluated against manually annotated segmentation of anatomical structures (referred to as ground-truth) in a 5-fold cross-validation strategy. The Dice Similarity Coefficient (DSC)[20] scores were calculated to measure the degree of overlap between predictions and the ground truth, where a 0 DSC corresponds to no overlap and 1 DSC indicates perfect overlap. While DSC is a standard segmentation metric for medical image segmentations,[21] its functional application to surgery can often be limited for task-specific requirements, where measuring distance to a structure is more relevant. Therefore, to place more emphasis on structure boundaries while still understanding prediction misalignment relative to the ground-truth, manual segmentations we also evaluated our DL framework with a surface boundary-based metric, Surface distance.

Average surface distance (ASD)[21] measures the shape-wise spatial error (in millimeters) between predictions and ground-truth segmentation. For computation of ASD, the paired predicted segmentation and ground-truth segmentation were first converted to 3D point clouds. Then, the closest distance for each point in the predicted and the ground truth segmentation was calculated. Finally, the average of these bidirectional distances from each point is used as a quantitative measure. For qualitative spatial visualization, the individual error of each point to the ground truth was assigned to the prediction segmentation mesh, resulting in the heat map that denotes the location of largest error (see **Figure 3**). To obtain a comprehensive assessment across all ground-truth annotations, we utilized mean (standard deviation) as well as median (interquartile range) to depict the typical accuracy of the model in the presence of extreme cases (ie, best and worst-performing scans).

DL Runtime Analysis

The total time span of automated segmentation pipeline was recorded, including critical steps of the DL framework and key components of the workflow. Differences between time stamps were calculated to determine the runtime of each corresponding step. The data preprocessing and training were performed on the AMD Ryzen Threadripper 3970 × 32-Core CPU and Nvidia RTX3090 GPU. The average execution time of image preprocessing is about 25 minutes and DL training is about 2 hours with 250 epochs for each fold. Notably, the training time remained consistent irrespective of the number of scans or labels used. Additionally, the model's inference time per scan was found to be under 5 seconds.

Results

The DL framework was validated across varying numbers of labeled scans and two distinct task settings: training on large anatomical structures (including NS, IT, and MS) and training on all anatomical structures (NS, IT, MS, and ON). These experiments shed light on the influence of labeled scans and the inclusion of smaller structures on the DL framework's performance.

Quantitative Analysis

The performance of the DL framework across large structures that utilized a varying number of labeled scans is listed in **Tables 1** and **2**. Empirical results demonstrate that the model can generate sub-mm ASDs with only four labeled scans. The performance of both the segmentation

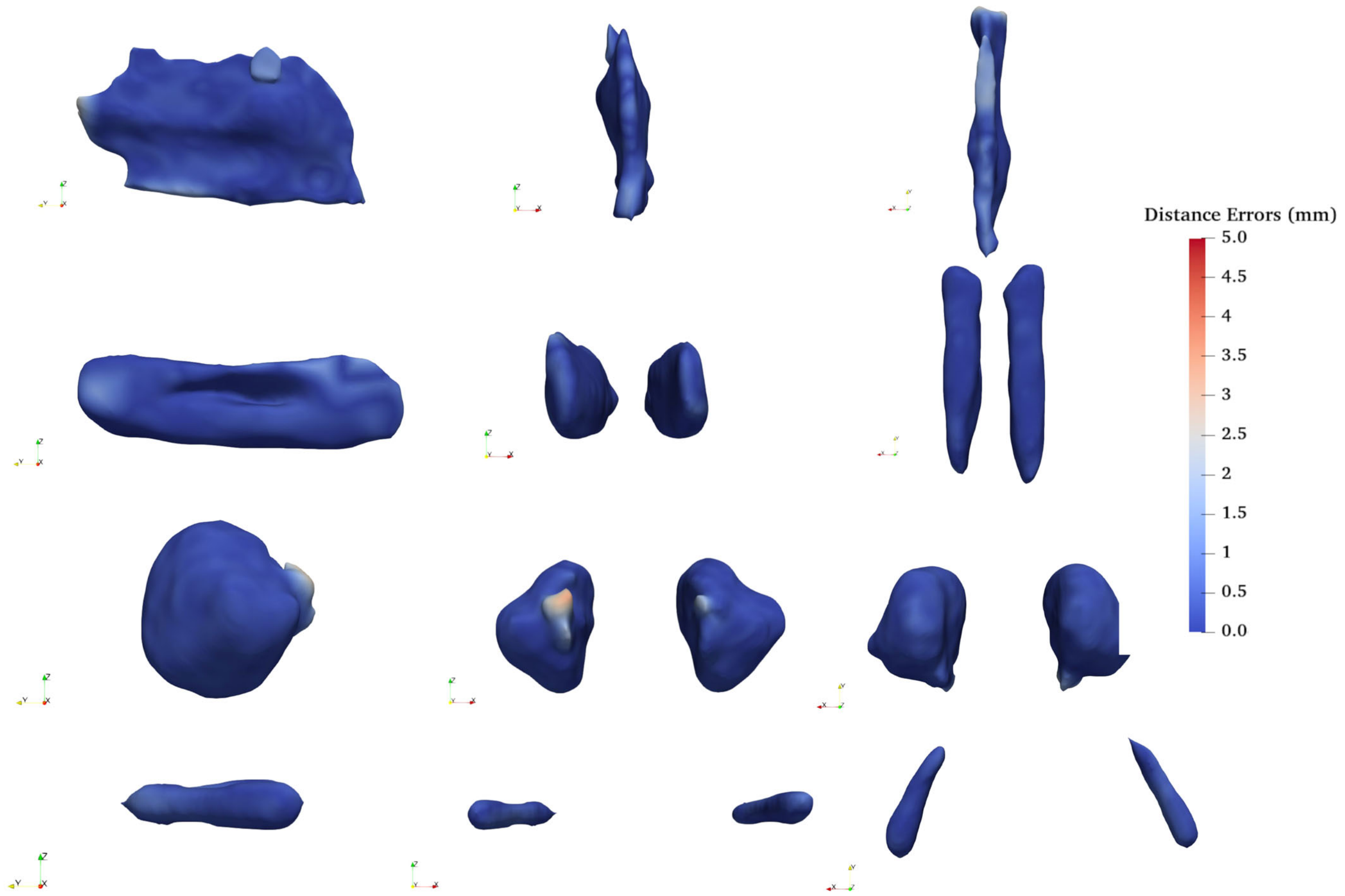

Figure 3. Qualitative visualization of the three structures (from top to bottom: NS, IT, MS, and ON) where each row shows side (left), front (middle), and lateral (right) view of heat map.

Table 1. Performance Comparison of the Registration Model Using Mean (std) for DSC and ASD (in mm) Computed Between Predicted and Ground-Truth Segmentations Across Anatomical Structures

	Registration network							
	1 Label		2 Labels		4 Labels		8 Labels	
Labels	DSC	ASD	DSC	ASD	DSC	ASD	DSC	ASD
NS	0.63 (0.07)	1.17 (0.21)	0.69 (0.03)	1.00 (0.11)	0.73 (0.06)	0.91 (0.16)	0.72 (0.08)	0.92 (0.24)
IT	0.67 (0.02)	1.20 (0.10)	0.77 (0.07)	0.85 (0.21)	0.78 (0.03)	0.84 (0.14)	0.77 (0.06)	0.93 (0.15)
MS	0.86 (0.06)	0.99 (0.47)	0.89 (0.05)	0.82 (0.35)	0.89 (0.04)	0.80 (0.34)	0.92 (0.04)	0.70 (0.28)

DSC value closer to 1 depicts higher overlap with the manual segmentations, while ASD close to 0 depicts higher agreement with the ground-truth shape. Abbreviations: ASD, average surface distance; DSC, Dice Similarity Coefficient; IT, inferior turbinate; MS, maxillary sinus; NS, nasal septum.

and the registration model is comparable, thereby indicating that both models were able to mutually guide each other during training to reach an optimal state for both models. Additionally, the performance of the model increases as the number of training labels increases. It is important to note that the performance trend is consistent across all the sinonasal structures, highlighting the generalizability of the two models.

Tables 3 and **4** present the performance of the DL framework across all structures. The results indicate that the proposed method can effectively generalize for both larger and smaller, critical structures essential for surgical scenarios. Additionally, the inclusion of additional structures, such as the optical nerve (ON), minimally affects the prediction accuracy for other structures, suggesting scalability of the approach. The empirical results suggest that the inclusion of smaller structures (eg, ON) has limited impact on the prediction for larger structures. Furthermore, the model's performance continues to improve with the number of available labeled scans. A violin plot depicted in **Figure 4** provides further insights, indicating that only four labeled scans are necessary to achieve median sub-mm

Table 2. Performance Comparison of the Segmentation Model Using Mean (std) for DSC and ASD (in mm) Computed Between Predicted and Ground-Truth Segmentations Across Anatomical Structures

	Segmentation network							
	1 Label		2 Labels		4 Labels		8 Labels	
Labels	DSC	ASD	DSC	ASD	DSC	ASD	DSC	ASD
NS	0.69 (0.09)	1.32 (0.34)	0.81 (0.04)	0.93 (0.26)	0.84 (0.01)	0.75 (0.12)	0.84 (0.03)	0.62 (0.17)
IT	0.72 (0.02)	1.22 (0.07)	0.86 (0.04)	0.70 (0.24)	0.88 (0.02)	0.58 (0.16)	0.89 (0.04)	0.52 (0.18)
MS	0.88 (0.03)	2.19 (1.11)	0.92 (0.03)	1.15 (0.74)	0.93 (0.02)	1.14 (0.43)	0.94 (0.04)	0.48 (0.27)

DSC value closer to 1 depicts higher overlap with the manual segmentations, while ASD close to 0 depicts higher agreement with the ground truth shape.
Abbreviations: ASD, average surface distance; DSC, Dice Similarity Coefficient; IT, inferior turbinate; MS, maxillary sinus; NS, nasal septum.

Table 3. Performance Comparison of the Segmentation Model Using Mean (Std) for DSC and ASD (in Mm) Computed Between Predicted and Ground-Truth Segmentations Across Anatomical Structures

	Segmentation network							
	1 Label		2 Labels		4 Labels		8 Labels	
Labels	DSC	ASD	DSC	ASD	DSC	ASD	DSC	ASD
NS	0.61 (0.06)	1.60 (0.29)	0.57 (0.13)	1.62 (0.43)	0.77 (0.04)	0.97 (0.19)	0.83 (0.02)	0.72 (0.16)
IT	0.65 (0.06)	1.77 (0.41)	0.69 (0.10)	1.52 (0.55)	0.84 (0.06)	0.80 (0.25)	0.88 (0.03)	0.57 (0.17)
MS	0.92 (0.03)	0.63 (0.21)	0.85 (0.16)	0.97 (0.87)	0.88 (0.14)	0.75 (0.66)	0.95 (0.02)	0.45 (0.21)
ON	0.21 (0.15)	4.73 (5.04)	0.27 (0.20)	2.91 (1.87)	0.47 (0.15)	1.47 (0.70)	0.60 (0.14)	1.12 (0.64)

DSC value closer to 1 depicts higher overlap with the manual segmentations, while ASD close to 0 depicts higher agreement with the ground-truth shape.
Abbreviations: ASD, average surface distance; DSC, Dice Similarity Coefficient; IT, inferior turbinate; MS, maxillary sinus; NS, nasal septum.

Table 4. Performance Comparison of the Segmentation Model Using Median (Interquartile Range) for DSC and ASD (in mm) Computed Between Predicted and Ground-Truth Segmentations Across Anatomical Structures

	Segmentation network							
	1 Label		2 Labels		4 Labels		8 Labels	
Labels	DSC	ASD	DSC	ASD	DSC	ASD	DSC	ASD
NS	0.61 (0.04)	1.56 (0.28)	0.61 (0.12)	1.55 (0.41)	0.77 (0.05)	0.96 (0.3)	0.83 (0.03)	0.72 (0.16)
IT	0.66 (0.09)	1.68 (0.54)	0.69 (0.18)	1.44 (0.6)	0.84 (0.05)	0.73 (0.25)	0.88 (0.03)	0.56 (0.23)
MS	0.92 (0.03)	0.58 (0.33)	0.92 (0.06)	0.65 (0.45)	0.94 (0.08)	0.43 (0.57)	0.95 (0.01)	0.38 (0.09)
ON	0.16 (0.25)	3.05 (2.16)	0.27 (0.34)	2.25 (2.77)	0.5 (0.16)	1.18 (0.59)	0.65 (0.15)	0.8 (0.33)

DSC value closer to 1 depicts higher overlap with the manual segmentations, while ASD close to 0 depicts higher agreement with the ground-truth shape.
Abbreviations: ASD, average surface distance; DSC, Dice Similarity Coefficient; IT, inferior turbinate; MS, maxillary sinus; NS, nasal septum.

ASDs for large sinonasal structures (NS: 0.96 mm, IT: 0.74 mm, MS: 0.43 mm) across the entire dataset, while eight scans are required for smaller structures like ON (0.8 mm).

Qualitative Analysis

Heatmap visualizations of predicted segmentations compared to the ground truth manual segmentations are shown in **Figure 3**. It shows that the proposed framework preserves the anatomical structures and their boundaries with minimal errors. The discrepancies occur primarily at the anterior aspect of the nasal septum which is a thin plate of bone and cartilage compared to the 3D shapes of the IT and MS. In a clinical context, anterior errors are more tolerable than posterior errors as most surgical work occurs deeper in the nasal cavity. However, such errors do limit the generalizability of the DL framework to structures that are located exclusively within the operative field or are thin. Future iterations of the DL framework will improve segmentation consistency across anatomical structures making this concern less relevant.

Discussion

Accurate segmentation of key anatomical structures within patient scans is a prerequisite for preoperative

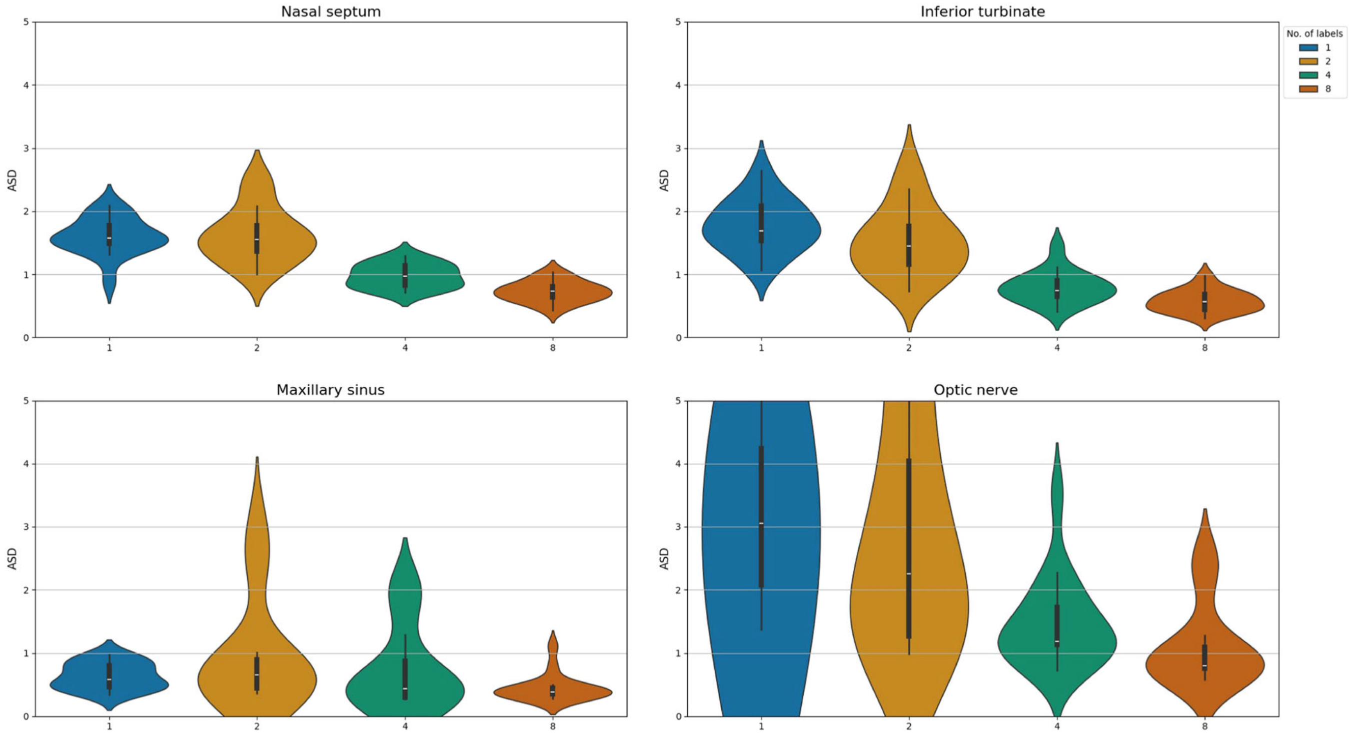

Figure 4. Plot depicting the performance of the segmentation model versus the number of labeled scans used for training the deep learning (DL) framework—DSC (left) and average surface distance (ASD) (right). Violin plots illustrate the distribution of ASD scores for our method. Each "violin" represents the accuracy of the model's predictions across whole dataset. The wider parts of the violins show where most of the ASD values lie. The white dot inside each violin represents the median accuracy, while the black vertical lines extending from the violins indicate the range of typical scores. The shape and spread of these violins across different label subgroups show that the performance of the model increases with the number of labeled scans. The median score (white dot) across different label subgroups shows that only four labeled scans are necessary to achieve sub-mm ASDs for large sinonasal structures—NS (0.96 mm), IT (0.74 mm), and MS (0.43 mm), whereas eight scans are required for critical structures—ON (0.8 mm).

planning workflows and development of patient-specific anatomical models for image- or robot-guided navigation systems. However, obtaining high-quality manual segmentation of relevant structures is difficult and time-consuming.[22] In this study, it took an experienced labeler about 15–20 hours to annotate three structures within 10 scans in a head-neck data set using *3D Slicer* while the well-built DL model may only take a few seconds to make automated inferences for each scan.

Despite the recent success of DL-based methods for automated segmentation of anatomical structures, a key drawback of such systems is the need for sufficiently large, manual-labeled datasets to develop robust models. In contrast, the *proposed* method requires the users to provide a much smaller sample of labeled segmentations and a pool of unlabeled scans to facilitate effective training of DL models. Empirical results demonstrate that even four labels are sufficient to obtain sub-mm performance.

Despite being label-efficient, a limitation of our framework is the processing requirement to initially coregister images. While image coregistration ensures consistent geometry and voxel spacing, it can lead to loss of information near anatomical boundaries. Another limitation of our framework is the computational resource and time requirements for the joint training of the segmentation/registration model. While the model's inference time is within 5 seconds, the training requires 2 hours.

While the current segmentation pipeline has the potential to facilitate preoperative planning, the utility for intraoperative image guidance is determined by the combined errors of anatomical segmentation and endoscope-to-patient registration. This study has demonstrated a crucial step towards the first of these two errors by generating accurate volumetric segmentation of patient anatomy on CT imaging.

Conclusion

This study presented a novel platform for automatically segmenting structures of the nasal cavity achieving sub-millimeter accuracy. Unlike conventional semi-automated atlas-based methods and resource-intensive DL-based segmentation approaches, this automated framework requires a lower number of manually segmented ground truths, thereby showcasing its utility in a clinical setting where limited time and resources preclude extensive manual labeling. This framework can be utilized for automation of pre-operative planning workflows, as well as construction of statistical shape models to quantify population variations. Furthermore, this pipeline has the potential to interface with image-guidance navigation systems and could therefore assist in understanding the location of critical anatomical structures during surgery. Future work will focus on evaluating the adaptability of this pipeline to various sinonasal structures, datasets, and imaging modalities, including MRI. The ongoing development of methods capable of accurate segmentation in a label-efficient manner holds promise for facilitating large-scale studies and establishing a foundation for effective intra-operative surgical guidance.

Author Contributions

Manish Sahu, conception/design, data acquisition, data analysis, manuscript drafting, data presentation, final approval; **Yuliang Xiao**, conception/design, data acquisition, data analysis, manuscript drafting, data presentation, final approval. **Jose L. Porras**, conception/design, data analysis, manuscript drafting, final approval; **Ameen Amanian**, conception/design, data acquisition, data analysis, manuscript drafting, final approval; **Aseem Jain**, conception/design, manuscript drafting, final approval; **Andrew Thamboo**, conception/design, manuscript drafting, final approval; **Francis X. Creighton**, conception/design, manuscript drafting, final approval; **Russell H. Taylor**, conception/design, manuscript drafting, final approval; **Masaru Ishii**, conception/design, data acquisition, manuscript drafting, final approval.

Disclosure

Competing interests: None.

Funding source: None.

ORCID iD

Manish Sahu http://orcid.org/0000-0003-2158-823X

References

1. Rosenfeld RM, Piccirillo JF, Chandrasekhar SS, et al. Clinical practice guideline (update): adult sinusitis. *Otolaryngol Head Neck Surg*. 2015;152(2 Suppl):S1-S39. doi:10.1177/0194599815572097
2. Labruzzo SV, Aygun N, Zinreich SJ. Imaging of the paranasal sinuses. *Otolaryngol Clin North Am*. 2015;48(5):805-815. doi:10.1016/j.otc.2015.05.008
3. Strauss G, Limpert E, Strauss M. Untersuchungen zur Effizienz eines Navigationssystems für die HNO-Chirurgie: Auswertungen von 300 Patienten [Evaluation of a daily used navigation system for FESS]. *Laryngorhinootologie*. 2009;88(12):776-781. doi:10.1055/s-0029-1237352
4. Lorenz KJ, Frühwald S, Maier H. Einsatz des Brainlab-Kolibri-Navigationssystems bei der endoskopischen Nasennebenhöhlenchirurgie in Lokalanästhesie. Erfahrung an 35 Patienten [The use of the Brain LAB Kolibri navigation system in endoscopic paranasal sinus surgery under local anaesthesia. an analysis of 35 cases]. *HNO*. 2006;54(11):851-860. doi:10.1007/s00106-006-1386-7
5. Linxweiler M, Pillong L, Kopanja D, et al. Augmented reality-enhanced navigation in endoscopic sinus surgery: a prospective, randomized, controlled clinical trial. *Laryngoscope Investig Otolaryngol*. 2020;5(4):621-629.
6. Leonard S, Reiter A, Sinha A, Ishii M, Taylor R, Hager G. Image-based navigation for functional endoscopic sinus surgery using structure from motion. *Proc SPIE*. 2016;9784:97840V.
7. Vrtovec T, Močnik D, Strojan P, Pernuš F, Ibragimov B. Auto-segmentation of organs at risk for head and neck radiotherapy planning: from atlas-based to deep learning methods. *Med Phys*. 2020;47(9):e929-e950.
8. 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*. 2019;38(1):99-106.
9. Konuthula N, Perez FA, Maga AM, et al. Automated atlas-based segmentation for skull base surgical planning. *Int J Comput Assist Radiol Surg*. 2021;16(6):933-941.
10. Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29(1):196-205.
11. Hoang Duc AK, Eminowicz G, Mendes R, et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. *Med Phys*. 2015;42(9):5027-5034.
12. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211.
13. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.
14. Ding AS, Lu A, Li Z, et al. A self-configuring deep learning network for segmentation of temporal bone anatomy in cone-beam CT imaging. *Otolaryngol Head Neck Surg*. 2023.
15. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012;30(9):1323-1341.
16. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011;54(3):2033-2044.
17. Xu Z, Niethammer M. DeepAtlas: joint semi-supervised learning of image registration and segmentation. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference. Shenzhen, China. October 13–17, 2019, Proceedings, Part II 22: Springer International Publishing; 2019:420-429.
18. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*. Munich, Germany. October 5–9,

2015, Proceedings, Part III 18: Springer International Publishing; 2015:234-241.

19. Silversmith W. *cc3d: Connected components on multilabel 3d images*. Zenodo; 2021. https://zenodo.org/record/5535251
20. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index1. *Academic Radiol*. 2004;11(2):178-189.
21. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15(1):29.
22. Ding AS, Lu A, Li Z, et al. Automated registration-based temporal bone computed tomography segmentation for applications in neurotologic surgery. *Otolaryngol Head Neck Surg*. 2022;167(1):133-140.