Journal Pre-proof

Automatic segmentation of the pharyngeal airway space with convolutional neural network

Sohaib Shujaat , Omid Jazil , Holger Willems ,
Adriaan Van Gerven , Eman Shaheen , Constantinus Politis ,
Reinhilde Jacobs

PII: S0300-5712(21)00126-3
DOI: <https://doi.org/10.1016/j.jdent.2021.103705>
Reference: JJOD 103705

To appear in: *Journal of Dentistry*

Received date: 10 February 2021
Revised date: 30 April 2021
Accepted date: 24 May 2021

Please cite this article as: Sohaib Shujaat , Omid Jazil , Holger Willems , Adriaan Van Gerven , Eman Shaheen , Constantinus Politis , Reinhilde Jacobs , Automatic segmentation of the pharyngeal airway space with convolutional neural network, *Journal of Dentistry* (2021), doi: <https://doi.org/10.1016/j.jdent.2021.103705>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Ltd.

Automatic segmentation of the pharyngeal airway space with convolutional neural network

Short title: Automatic upper airway segmentation

Sohaib Shujaat^{1*}, Omid Jazil¹, Holger Willems², Adriaan Van Gerven², Eman Shaheen¹,
Constantinus Politis¹, Reinhilde Jacobs^{1,3}

^{1*}OMFS IMPATH Research Group, Department of Imaging & Pathology, Faculty of
Medicine, KU Leuven & Oral and Maxillofacial Surgery, University Hospitals Leuven,
Leuven, Belgium

²Relu, R&D, Leuven, Belgium,

³Department of Dental Medicine, Karolinska Institutet, Stockholm, Sweden

Corresponding author:

Sohaib Shujaat, OMFSIMPATH Research Group,

Campus Sint-Rafaël, Kapucijnenvoer 33, BE-3000 Leuven, Belgium

Email: sohaib.shujaat941@gmail.com

Tel no.: +32488065857

Key words: computer neural networks, pharynx, deep learning, three-dimensional imaging

Automatic segmentation of the pharyngeal airway space with convolutional neural network

Journal Pre-proof

Abstract

Objectives

This study proposed and investigated the performance of a deep learning based three-dimensional (3D) convolutional neural network (CNN) model for automatic segmentation of the pharyngeal airway space (PAS).

Methods

A dataset of 103 computed tomography (CT) and cone-beam CT (CBCT) scans was acquired from an orthognathic surgery patients database. The acquisition devices consisted of 1 CT (128-slice multi-slice spiral CT, Siemens Somatom Definition Flash, Siemens AG, Erlangen, Germany) and 2 CBCT devices (Promax 3D Max, Planmeca, Helsinki, Finland and Newtom VGi evo, Cefla, Imola, Italy) with different scanning parameters. A 3D CNN-based model (3D U-Net) was built for automatic segmentation of the PAS. The complete CT/CBCT dataset was split into three sets, training set (n=48) for training the model based on the ground-truth observer-based manual segmentation, test set (n=25) for getting the final performance of the model and validation set (n=30) for evaluating the model's performance versus observer-based segmentation.

Results

The CNN model was able to identify the segmented region with optimal precision (0.97 ± 0.01) and recall (0.96 ± 0.03). The maximal difference between the automatic segmentation and ground truth based on 95% hausdorff distance score was 0.98 ± 0.74 mm. The dice score of 0.97 ± 0.02 confirmed the high similarity of the segmented region to the ground truth.. The Intersection over union (IoU) metric was also found to be high (0.93 ± 0.03). Based on the acquisition devices, Newtom VGi evo CBCT showed improved performance compared to the Promax 3D Max and CT device.

Conclusion

The proposed 3D U-Net model offered an accurate and time-efficient method for the segmentation of PAS from CT/CBCT images.

Clinical significance

The proposed method can allow clinicians to accurately and efficiently diagnose, plan treatment and follow-up patients with dento-skeletal deformities and obstructive sleep apnea which might influence the upper airway space, thereby further improving patient care.

Key words: computer neural networks, pharynx, deep learning, three-dimensional imaging

1. Introduction

The pharyngeal airway space (PAS), also referred to as the upper airway, is a complex anatomical region having an intimate relationship with the surrounding skeletal and soft tissue structures. It is mainly responsible for performing functions such as respiration, phonation and swallowing [1]. The PAS assessment has been an area of interest for clinicians since several studies confirmed its relationship with the craniofacial growth and development [[2], [3], [4], [5]]. Previously, two-dimensional (2D) lateral cephalometry was applied for assessing the airway changes at the stages of diagnosis, treatment planning and follow-up in patients with dento-facial and skeletal abnormalities[6,7]. However, based on the potential limitations offered by the 2D techniques, these have been widely replaced by computed tomography (CT), cone-beam CT (CBCT) and magnetic resonance imaging (MRI) as a clinical standard for analyzing PAS volume and dimensional changes for a better understanding of its pathophysiology [[8], [9], [10]].

The first and the most vital step for analyzing PAS volume consists of segmentation which enables the delineation and separation of the airway space from the rest of the scan, thereby, allowing 3D visualization and quantification. Over the past decade, various upper airway segmentation tools and algorithms have been developed which are either manual, semi-automatic or automatic in nature [11]. Although manual segmentation does offer the most accurate replication of the anatomical structure and is considered as a gold standard, nevertheless, it is time-consuming and labor-intensive. Various threshold-based semi-automatic software packages have been validated for volumetric assessment as well, where the user defines a volume of interest (VOI) and the software automatically combines the gray threshold values in that region without efficiently considering the image intensity and anatomical variations, thereby manual post-processing is required for corrections [12]. Similarly, some studies have suggested a fixed threshold value for segmenting PAS [13,14],

which might vary depending on the CBCT device, scanning parameters, machine calibration and the noise caused by the patient movement or metal artefacts[15,16]. Studies have also proposed fully automatic sophisticated and hybrid image segmentation algorithms for the segmentation of the PAS. However they offer limited value based on either insufficient accuracy, fixed thresholding, localization of seed points through manual intervention, manual VOI selection, dependency on image orientation or failure of the algorithm across variable scanning parameters [[12][17][18][19]].

Recently, deep learning convolutional neural networks (CNNs), a multilayer structure-learning algorithm has gained much attention in the dentomaxillofacial field, allowing processing of the data that mimics a brain functioning through neural networks and automatic learning of complex data [20]. Hence, providing with an accurate voxel-wise segmentation. The CNNs have been successfully applied to segment medical image data from different acquisition devices [21]. Nevertheless, lack of evidence exists related to the automatic CT/CBCT based segmentation of the PAS by applying a deep learning based CNN model. This study lays the groundwork to overcome the shortcomings associated with previous methods for potential future investigations. Therefore, the aim of the study was to propose and investigate the performance of a deep learning based 3D CNN model for PAS segmentation from CT/CBCT images.

2. Materials and methods:

This study was conducted in compliance with the World Medical Association Declaration of Helsinki on medical research. Ethical approval was obtained from the Local Ethical Review Board and all data were anonymized.

2.1 Dataset

The total dataset consisted of 103 scans acquired from an orthognathic surgery patients database (LORTHOG), which varied across devices and scanning parameters. The acquisition

devices consisted of 1 CT and 2 CBCT devices: 128-slice multi-slice spiral CT (Siemens Somatom Definition Flash, Siemens AG, Erlangen, Germany, 120 kV, 161 mAs, slice thickness: 0.75mm), Promax 3D Max (Planmeca, Helsinki, Finland, 96 kV, 216 mAs, slice thickness: 0.6mm, field of view: $230 \times 260 \text{ mm}^2$) and Newtom VGi evo (Cefla, Imola, Italy, 110 kV, 15.3 mAs, slice thickness: 0.3, field of view: $240 \times 190 \text{ mm}^2$). The dataset included CT/CBCT scans of patients with dentofacial deformities and were randomly recruited at three phases, pre-orthognathic surgery, immediate postoperative phase (1–6 weeks) and 1-2 years postoperatively. Thereafter, the complete dataset was randomly split into three sets:

- i. Training set (n=48), a subset for training and fitting the CNN model with manually segmented data.
- ii. Test set (n=25), a subset for testing the model performance based on the trained data
- iii. Validation set (n=30), a subset used to evaluate the model performance by comparing it with manually segmented data.

The PAS training and validation ground-truth datasets were prepared by performing manual segmentation of the CT/CBCT images in a Digital Imaging and Communications in Medicine (DICOM) format using Mimics software (version 22.0, Materialise N.V., Leuven, Belgium). All images were oriented according to the Frankfurt horizontal plane. The PAS was limited by the nasal cavity superiorly, oral cavity anteriorly, pharyngeal border posteriorly and inferiorly till the limit of the scan which was mostly at the level of 2nd, 3rd or 4th cervical vertebrae depending on the field of view of the scan. Figure 1 illustrates examples of two cases showing the limits of segmentation. Two trained and calibrated experts segmented the PAS individually from different datasets using manual thresholding which was initially set between the range of -1024 to -500HU (Hounsfield Unit) to create a mask. Thereafter, the mask contours were edited using livewire, lasso and eclipse function for addition or removal of the

under/over-extensions and filling of the voids if present within the limits of the PAS.

Delineation was performed in axial, sagittal and coronal planes. Following that, the achieved mask was converted to a standard tessellation language (STL) format. All contours and 3D segmentations were viewed by both observers simultaneously to reconfirm the delineation of the PAS and alterations were carried out if any discrepancy existed. Later, both observers repeated the segmentation of 10 cases each blindly at an interval of 1 month for assessing the inter- and intra-observer reliability based on the retrieved volumetric data. The mean difference of these segmentations was further confirmed by applying a part comparison analysis in 3-matic medical software (version 14.0, Materialise, Leuven, Belgium).

2.2 AI segmentation framework

The AI models applied were 3D U-Nets [22]. The U-Nets in this work were trained with a Binary Cross Entropy loss: $L_{BCE} = y_n * \log(p_n) + (1 - y_n) * \log(1 - p_n)$, where for each voxel n with a ground truth value $y_n = 0$ or 1 and predicted probability of the network p_n .

The networks were trained using CT/CBCT DICOM images, and STL images were converted to binary masks with early stopping on the validation set to prevent overfitting. The early stopping criterion is a regularization process to decrease the generalization error by randomly further splitting the training set into training and validation sets and the performance of validation set is monitored during the training process. The validation set error normally decreases following initial training, however, at a certain point the error begins to rise with data overfitting which means that no additional knowledge is being learnt by the training set and at that stage to avoid the validation loss an early stopping is applied. The validation set in this step was a part of the training set and not the original validation set for which performance metrics were measured.

A two-step approach was applied for automatically segmenting the airway. In the first step, the full scan was resized to a fixed size at a lower resolution and was uploaded to a U-Net that

was trained to segment low resolution scans. In the second step, this low resolution approximation was used to select several full resolution patches. These patches were then transferred to another U-net trained to segment full resolution patches. The patch segmentations were then combined to create the final segmentation map. Figure 2 illustrates the complete workflow for segmenting the airway. The AI model was deployed and transferred on an online customized user-interactive cloud-based platform (version 1.0, Toothflow, Relu, Inc, Leuven. Belgium) allowing observers to upload the DICOM dataset, visualize and export the DICOM data and the 3D reconstructed PAS in a STL format. The Toothflow platform is a an interactive platform for artificial intelligence (AI)-driven segmentations of dentofacial anatomy. The platform creates an automatic segmentation of the dentofacial structure with AI and allows the specialist to further refine the details of the segmentation [23].

2.3 Validation metrics

A confusion matrix was applied for the voxel-wise comparison of the ground truth and CNN model prediction by utilizing four variables: true positive (TP), true negative (TN), false positive (FP) and false negative (FN) voxels. TP represented the voxels of the PAS which were supposed to be segmented and were correctly segmented, FN referred to the voxels which should have been segmented but were not segmented, FP were the voxels that got segmented even though they were not supposed to segment, and TN involved voxels which were not supposed to be segmented so they were not included in the final segmented region. Based on the aforementioned variables, the performance of the AI model was assessed according to the following performance measures:

- i. **Precision** showed percentage of the accurately identified segmented region out of the completely segmented region.

$$Precision = \frac{TP}{TP + FP}$$

- ii. **Recall** showed the rate of airway region which was perfectly detected.

$$Recall = \frac{TP}{TP + FN}$$

- iii. **Accuracy** referred to the rate of correct findings in relation to all the findings observed.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- iv. **Dice score** provided the score on how much the segmented region was similar to the ground truth.

$$Dice = \frac{2 * TP}{2 * TP + FP + FN}$$

- v. **Intersection over union (IoU)** represented the amount of overlap between the predicted segmentation and the ground truth.

$$IoU = \frac{TP}{TP + FP + FN}$$

- vi. **95% Hausdorff Distance (HD)** provided the maximal difference between the automatic segmentation and ground truth.

$$hd(P, G) = \max_{p \in P} \min_{g \in G} \|p - g\|_2$$

$$hd(G, P) = \max_{g \in G} \min_{p \in P} \|p - g\|_2$$

$$HD = \max\{hd(P, G), hd(G, P)\}$$

3. Statistical analysis

The data were analyzed with IBM SPSS Statistics for Windows, version 21.0 (IBM Corp., Armonk, NY, USA). Mean and standard deviations were calculated for each evaluation metric and error of manual segmentation., Intra-class correlation coefficient (ICC) was applied at a 95% confidence interval for calculating the inter- and intra-observer reliability of the manual segmentation. The performance metrics for each imaging modality were compared using one-way ANOVA followed by Tukey's multiple comparison. A p value of <0.05 was considered significant.

4. Results

The average time required for the data processing and segmentation with the CNN model was 14.5 ± 2.3 seconds and 70.0 ± 26.9 minutes when applying manual segmentation irrespective of the imaging modality. The ICC for both inter- and intra-observer manual segmentation was excellent. The intra-observer ICC was found to be 0.998 (95% confidence interval: 0.991-0.999) with a mean error of $0.06 \pm 0.26 \text{ mm}^3$, whereas the inter-observer reliability was 0.996 (95% confidence interval: 0.986-0.999) with a mean error of $0.05 \pm 0.30 \text{ mm}^3$. No significant difference was detected amongst observers ($p=0.9378$)

Table 1 demonstrates the overall and imaging modality based performance of the CNN model for segmenting PAS compared to the ground truth. The CNN model offered a high precision (0.97 ± 0.01) and recall (0.96 ± 0.03), indicating an accurate and near to perfect segmentation of the PAS. When comparing the overlap of the ground truth with AI model, the 95% HD score revealed a value of $0.98 \pm 0.74 \text{ mm}$, suggesting towards an optimal overlap. Figure 3 illustrates an example of a case showing an overlap between the CNN predicted model and ground truth. Additionally the model offered a 100% accuracy for correctly segmenting the airway. The high dice score (0.97 ± 0.02) confirmed that the automatically segmented region was able to well reflect the manually segmented PAS. The mean IOU of 0.93 ± 0.03 also confirmed an optimal spatial overlap in comparison to the ground truth.

When comparing the dataset based on imaging modality, Newtom VGi evo CBCT showed the highest performance metrics, whereas, the multi-slice spiral CT had the lowest performance metrics amongst the 3 acquisition devices. Newtom CBCT device provided a significantly better scoring for recall ($p=0.0011$), dice ($p=0.0391$) and IOU ($p=0.0341$) compared to Promax 3D Max. At the same instance, offering a significantly improved recall ($p=0.0251$) and 95% HD ($p=0.0211$) compared to the CT device. When comparing Promax 3D Max with the CT device, only the 95% HD performance parameter was found to offer a significantly

improved score ($p=0.0603$) in favor of the CBCT device. Additionally the precision of the AI model was not significantly influenced by the type of imaging modality where all modalities offered an equal accuracy.

5. Discussion

In dentomaxillofacial surgery, the most important aspect from the patient treatment planning, follow-up evaluation and research perspective is to achieve a precise and fast segmentation of the anatomical structures. Manual delineation still remains the gold standard for an accurate PAS segmentation followed by semi- and fully automatic software programs, as currently no other alternative exists offering comparable accuracy to that of manual delineation by an expert [24]. Some studies have suggested printing a phantom of known geometry to be applied as a gold standard to assess the accuracy of segmentation [12,25]. However, these phantoms fail to replicate the tortuous airway anatomy and to generate a clinically-oriented scenario with artefacts and image noise. Additionally, the error induced from the printing process and post-processing also cannot be ignored. The existing evidence lacks an efficient and reliable method for PAS segmentation applicable across different acquisition devices considering the image noise level and intensity inhomogeneities. In the light of the current technological advancements in AI, the current study focused towards providing an accurate and time-efficient deep learning CNN for segmentation of the PAS.

The proposed CNN architecture showed a remarkable performance for segmenting the PAS. A comparison with existing studies was deemed difficult due to the variability of the image acquisition device and methodologies mostly consisted of non-CNN based algorithms. Cheng et al. [26] proposed algorithm based pipeline for extracting upper airway from CT and CBCT images, however failed to either validate or provide accuracy metrics. Neelapu et al. [19] applied a knowledge-based algorithm to automatically segment the PAS from CBCT dataset and compared five segmentation techniques following automatic contour initialization. They

observed that the accuracy, precision and recall with all the segmentation methods mostly ranged between the score of 0.80 and 0.90. Salerno et al. [27] proposed a semi-automatic dynamic region-growing procedure to segment the PAS from CT scans of syndromic patients offering a high dice score of 0.94, which could have resulted due to the small validation data of only four patients. Recently, Sin et al. [28] proposed a CNN based algorithm for automatically segmenting the PAS from CBCT images with a high similarity rate compared to the observer, however, the authors relied on a semi-automatic region-growing segmentation process to create the ground truth. Although region-growing is faster than other approaches, but it requires user-interaction for placing a seed which can lead to over/under-segmentation of the region of interest depending on the locality where the seed is placed and is sensitive to image noise and contrast which causes leakage into the surrounding structures [29]. Ivanovska et al. [30] applied automatic intensity clustering algorithm for segmenting PAS from MRI images and achieved a dice score of 0.85 ± 0.31 , where manual initialization was performed by defining the slice which could have reduced the accuracy. In comparison to the aforementioned studies, our approach offered a fully automatic approach with higher performance and the metrics scoring fell within the range of 0.95-1.00. Also, it should be noted that the accuracy metric was calculated to allow comparison with other studies. However, when assessing segmentation performance, this metric should be avoided because it uses TN in the nominator, which lead to skewness in the data. As in our study the data skewed towards 100% accuracy. This skewness is known as accuracy paradox, where the high accuracy does not necessarily indicate high performance of the network and even a model with lower accuracy can outperform the model with higher accuracy. The is caused by the imbalance between the positive and negative classes which might lead to the over-estimation of the accuracy. Thereby, making accuracy an untrustworthy parameter for assessing the performance of a model [31]. To overcome these paradoxical findings associated with

accuracy, other metrics assessed in the study provide a more meaningful output in relation to the performance of the CNN model.

Another important aspect which has received less attention in literature has been the computational time of the segmentation methods. Our model allowed segmentation without any human intervention for VOI selection or manual correction in 14.5 ± 2.3 seconds.

Although, the current state-of-art semi-automatic commercial or open-source software programs offering PAS segmentation which usually rely on global thresholding are more time-efficient [24], with the processing time of around 1-2 seconds. Nevertheless, the global threshold can either over- or under-segment the structures and the manual anatomical modelling for accurately contouring is often indispensable [16]. In contrast, the proposed methodology offered a time-efficient network which optimally delineated and defined the boundaries of the PAS automatically without any need for post-processing.

Additionally, the accuracy of PAS segmentation may also be influenced by the type of the CT/CBCT scanning device, parameters and presence of noise or artefacts, which have also been neglected in the previous studies. All those factors affect the voxel values of the scan and the previously proposed algorithms failed to validate based on the different noise levels and intensity inhomogeneities [32]. The variation in image acquisition protocols have also been known to affect the deep learning algorithms and similar limitation exists for all auto-segmentation techniques [33]. Meanwhile, the training dataset of the current study offered more generalizability with the involvement of multivendor acquisition devices having different scanning parameters and presence of metal and/or motion artefacts. From the radiation protection point of view and taking into account the ethical considerations, the CT/CBCT images acquired in this study were recruited from an orthognathic surgery database, utilizing an ethically approved standardized protocol and scanning parameters as previously reported by Stratis et al [34]. These scanning parameters allowed the maxillofacial

surgeons to carefully assess the 3D anatomy and segment hard and soft tissue structures optimally during the treatment planning phase with a low dose scanning protocol. The aforementioned protocol is justified from both clinical and research perspective.

In our study, the Newtom CBCT device offered the highest performance compared to the other two devices. This could be attributed to the small sample size of Promax 3D Max and multi-slice spiral CT scans for training which could have influenced the CNN performance. Therefore, increasing the sample size of other devices training dataset could further enhance the performance. However, considering the fact that there is a wide range of CT and CBCT devices in the market and hospitals offering variable scanning parameters and image qualities with or without artefact reduction protocol [35,36], so gathering training dataset would be troublesome requiring a multi-center data collection. However, a probable solution to overcome this limitation might be to apply simulations based on various scanning conditions for further training of the model.

The study was associated with certain limitations. Firstly, the CNN only allowed segmentation of the PAS without the inclusion of sino-nasal airway sub-regions. Further research would focus on improving the algorithm and enable it to identify the pharyngeal and sino-nasal airway sub-regions with optimal performance. Secondly, the CNN only allowed visualization and export of the segmented region without any provision of dimensional and morphological measurements, thereby requiring further fine-tuning to include quantitative measures. Thirdly, as the CNN model was trained and validated with 3 acquisition devices, its further generalizability needs to be evaluated with other CT/CBCT devices.

6. Conclusions

We established a self-constrained deep learning based 3D U-Net model with encouraging performance for the detection and segmentation of PAS, thus providing clinicians and researchers with a time-efficient and labor-free method which can further increase their

diagnostic efficiency. Future studies should investigate the PAS segmentation with a larger training set offering anatomical variability and to compare the performance of different CNN architectures.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT author statement

Sohaib Shujaat: Conceptualization, Methodology, Software, Investigation, Formal analysis, Visualization, Investigation, Writing - Original Draft, Writing - Review & Editing

Omid Jazil: Validation, Investigation, Software, , Formal analysis

Eman Shaheen: Conceptualization, Supervision, Writing - Review & Editing

Holger Willems: Conceptualization, Methodology, Investigation, Software, Writing - Review & Editing

Adriaan Van Gerven: Conceptualization, Methodology, Investigation, Software, Writing - Review & Editing

Constantinus Politis: Conceptualization, Supervision, Writing - Review & Editing

Reinhilde Jacobs: Conceptualization, Visualization, Supervision, Writing - Review & Editing

References

- [1] G. Di Carlo, A. Polimeni, B. Melsen, P.M. Cattaneo, The relationship between upper airways and craniofacial morphology studied in 3D. A CBCT study, *Orthod. Craniofac. Res.* 18 (2015) 1–11. doi:10.1111/ocr.12053.
- [2] S.A. Schendel, R. Jacobson, S. Khalessi, Airway growth and development: A computerized 3-dimensional analysis, *J. Oral Maxillofac. Surg.* 70 (2012) 2174–2183. doi:10.1016/j.joms.2011.10.013.
- [3] L.V. Claudino, C.T. Mattos, A.C.D.O. Ruellas, E.F. Sant Anna, Pharyngeal airway characterization in adolescents related to facial skeletal pattern: A preliminary study, *Am. J. Orthod. Dentofac. Orthop.* 143 (2013) 799–809. doi:10.1016/j.ajodo.2013.01.015.

- [4] Z.H. Zheng, T. Yamaguchi, A. Kurihara, H.F. Li, K. Maki, Three-dimensional evaluation of upper airway in patients with different anteroposterior skeletal patterns, *Orthod. Craniofac. Res.* 17 (2014) 38–48. doi:10.1111/ocr.12029.
- [5] M. Celikoglu, M. Bayram, A.E. Sekerci, S.K. Buyuk, E. Toy, Comparison of pharyngeal airway volume among different vertical skeletal patterns: A cone-beam computed tomography study, *Angle Orthod.* 84 (2014) 782–787. doi:10.2319/101013-748.1.
- [6] V.A. Pereira-Filho, L.M. Castro-Silva, M. De Moraes, M.F.R. Gabrielli, J.A.D.B. Campos, P. Juergens, Cephalometric evaluation of pharyngeal airway space changes in class III patients undergoing orthognathic surgery, *J. Oral Maxillofac. Surg.* 69 (2011) e409–e415. doi:10.1016/j.joms.2011.02.132.
- [7] A.Y. Gungor, H. Turkkahraman, H.H. Yilmaz, M. Yariktas, Cephalometric comparison of obstructive sleep apnea patients and healthy controls, *Eur. J. Dent.* 7 (2013) 48–54. doi:10.1055/s-0039-1698995.
- [8] C. Aboudara, I. Nielsen, J.C. Huang, K. Maki, A.J. Miller, D. Hatcher, Comparison of airway space with conventional lateral headfilms and 3-dimensional reconstruction from cone-beam computed tomography, *Am. J. Orthod. Dentofac. Orthop.* 135 (2009) 468–479. doi:10.1016/j.ajodo.2007.04.043.
- [9] J. He, Y. Wang, H. Hu, Q. Liao, W. Zhang, X. Xiang, X. Fan, Impact on the upper airway space of different types of orthognathic surgery for the correction of skeletal class III malocclusion: A systematic review and meta-analysis, *Int. J. Surg.* 38 (2017) 31–40. doi:10.1016/j.ijso.2016.12.033.
- [10] N.A. Alsufyani, M.L. Noga, M. Witmans, P.W. Major, Upper airway imaging in sleep-disordered breathing: role of cone-beam computed tomography, *Oral Radiol.* 33 (2017) 161–169. doi:10.1007/s11282-017-0280-1.

- [11] S. El Khateeb, Three-Dimensional Image Segmentation of Upper Airway by Cone Beam CT: A Review of Literature, *Egypt. Dent. J.* 66 (2020) 1527–1535. doi:10.21608/edj.2020.25972.1074.
- [12] Weissheimer, L.M.E. De Menezes, G.T. Sameshima, R. Enciso, J. Pham, D. Grauer, Imaging software accuracy for 3-dimensional analysis of the upper airway, *Am. J. Orthod. Dentofac. Orthop.* 142 (2012) 801–813. doi:10.1016/j.ajodo.2012.07.015.
- [13] M.G. Lenza, M.M. de, M. Dalstra, B. Melsen, P.M. Cattaneo, An analysis of different approaches to the assessment of upper airway morphology: A CBCT study, *Orthod. Craniofacial Res.* 13 (2010) 96–105. doi:10.1111/j.1601-6343.2010.01482.x.
- [14] M. Alves, C. Baratieri, C.T. Mattos, D. Brunetto, R.D.C. Fontes, J.R.L. Santos, A.C.D.O. Ruellas, Is the airway volume being correctly analyzed?, *Am. J. Orthod. Dentofac. Orthop.* 141 (2012) 657–661. doi:10.1016/j.ajodo.2011.11.019.
- [15] M. van Eijnatten, J. Koivisto, K. Karhu, T. Forouzanfar, J. Wolff, The impact of manual threshold selection in medical additive manufacturing, *Int. J. Comput. Assist. Radiol. Surg.* 12 (2017) 607–615. doi:10.1007/s11548-016-1490-4.
- [16] J. Minnema, M. van Eijnatten, W. Kouw, F. Diblen, A. Mendrik, J. Wolff, CT image segmentation of bone for medical additive manufacturing using a convolutional neural network, *Comput. Biol. Med.* 103 (2018) 130–139. doi:10.1016/j.combiomed.2018.10.012.
- [17] H. Shi, W.C. Scarfe, A.G. Farman, Upper airway segmentation and dimensions estimation from cone-beam CT image datasets, *Int. J. Comput. Assist. Radiol. Surg.* 1 (2006) 177–186. doi:10.1007/s11548-006-0050-8.
- [18] N.A. Alsufyani, A. Hess, M. Noga, N. Ray, M.A.Q. Al-Saleh, M.O. Lagravère, P.W. Major, New algorithm for semiautomatic segmentation of nasal cavity and pharyngeal

- airway in comparison with manual segmentation using cone-beam computed tomography, *Am. J. Orthod. Dentofac. Orthop.* 150 (2016) 703–712. doi:10.1016/j.ajodo.2016.06.024.
- [19] B.C. Neelapu, O.P. Kharbanda, V. Sardana, A. Gupta, S. Vasamsetti, R. Balachandran, S.S. Rana, H.K. Sardana, A pilot study for segmentation of pharyngeal and sino-nasal airway subregions by automatic contour initialization, *Int. J. Comput. Assist. Radiol. Surg.* 12 (2017) 1877–1893. doi:10.1007/s11548-017-1650-1.
- [20] J.J. Hwang, Y.H. Jung, B.H. Cho, M.S. Heo, An overview of deep learning in the field of dentistry, *Imaging Sci. Dent.* 49 (2019) 1–7. doi:10.5624/isd.2019.49.1.1.
- [21] S.M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, M.K. Khan, Medical Image Analysis using Convolutional Neural Networks: A Review, *J. Med. Syst.* 42 (2018) 226. doi:10.1007/s10916-018-1088-1.
- [22] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: Learning dense volumetric segmentation from sparse annotation, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 9901 LNCS (2016) 424–432. doi:10.1007/978-3-319-46723-8_49.
- [23] Relu - cloud-app, <https://toothflow.com/> (accessed February 6, 2021).
- [24] N.A. Alsufyani, C. Flores-Mir, P.W. Major, Three-dimensional segmentation of the upper airway using cone beam CT: A systematic review, *Dentomaxillofacial Radiol.* 41 (2012) 276–284. doi:10.1259/dmfr/79433138.
- [25] H. Chen, M. Van Eijnatten, J. Wolff, J. De Lange, P. F. Van Der Stelt, F. Lobbezoo, G. Aarab, Reliability and accuracy of three imaging software packages used for 3D analysis of the upper airway on cone beam computed tomography images. *Dentomaxillofac. Radiol.* 46 (2017) 20170043. doi: 10.1259/dmfr.20170043.

- [26] I. Cheng, S. Nilufar, C. Flores-Mir, A. Basu, Airway segmentation and measurement in CT images, in: Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc., Annu Int Conf IEEE Eng Med Biol Soc, 2007: pp. 795–799. doi:10.1109/IEMBS.2007.4352410.
- [27] S. Salerno, C. Gagliardo, S. Vitabile, C. Militello, G. La Tona, M. Giuffrè, A. Lo Casto, M. Midiri, Semi-automatic volumetric segmentation of the upper airways in patients with Pierre Robin Sequence, *Neuroradiol. J.* 27 (2014) 487–494. doi:10.15274/NRJ-2014-10067.
- [28] Ç. Sin , N. Akkaya, S. Aksoy , K. Orhan , U. Öz, A deep learning algorithm proposal to automatic pharyngeal airway detection and segmentation on CBCT images, *Orthod. Craniofac. Res.* (2021). doi: 10.1111/ocr.12480. Online ahead of print.
- [29] M.L.U.R. Shahid, T. Chitiboi, T. Ivanovska, V. Molchanov, H. Völzke, L. Linsen, Automatic MRI segmentation of para-pharyngeal fat pads using interactive visual feature space analysis for classification, *BMC Med. Imaging.* 17 (2017) 15. doi:10.1186/s12880-017-0179-7.
- [30] T. Ivanovska, J. Dober, R. Laqua, K. Hegenscheid, H. Völzke, Pharynx segmentation from MRI data for analysis of sleep related disorders, in: *Lect. Notes Comput. Sci.* (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer, Berlin, Heidelberg, 2013: pp. 20–29. doi:10.1007/978-3-642-41914-0_3.
- [31] F.J. Valverde-Albacete, C. Peláez-Moreno, 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox, *PLoS One.* 9 (2014) 84217. doi:10.1371/journal.pone.0084217.
- [32] B. Ibragimov, L. Xing, Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks:, *Med. Phys.* 44 (2017) 547–557. doi:10.1002/mp.12045.

- [33] Rizwan I Haque, J. Neubert, Deep learning approaches to biomedical image segmentation, *Informatics Med. Unlocked*. 18 (2020) 100297.
doi:10.1016/j.imu.2020.100297.
- [34] A. Stratis, G. Zhang, X. Lopez-Rendon, C. Politis, R. Hermans, R. Jacobs, R. Bogaerts, E. Shaheen, H. Bosmans, Two examples of indication specific radiation dose calculations in dental CBCT and Multidetector CT scanners, *Phys. Medica*. 41 (2017) 71–77. doi:10.1016/j.ejmp.2017.03.027.
- [35] A.A. Schegerer, H.D. Nagel, G. Stamm, G. Adam, G. Brix, Current CT practice in Germany: Results and implications of a nationwide survey, *Eur. J. Radiol*. 90 (2017) 114–128. doi:10.1016/j.ejrad.2017.02.021.
- [36] H. Gaêta-Araujo, T. Alzoubi, K. de Faria Vasconcelos, K. Orhan, R. Pauwels, J.W. Casselman, R. Jacobs, cone beam computed tomography in dentomaxillofacial radiology: A two-decade overview, *Dentomaxillofac Radiol*. 49 (2020).
doi:10.1259/DMFR.20200145.

Figure legends

Figure 1. Example of two cases showing limits of the pharyngeal airway space (PAS) segmentation, anteriorly: oral cavity and soft tissue contour of pharyngeal wall, posteriorly: soft tissue contour of the pharyngeal wall, superiorly: openings into the nasal cavity, inferiorly: complete length of the PAS between the anteroposterior soft tissue contours.

Figure 2. Workflow of the 3D U-Net based segmentation model

Figure 3. Overlap between predicted model (red color) and ground truth (green color) in axial, coronal and sagittal plane with 3D reconstruction

Table 1. Convolutional neural network's segmentation performance compared to ground-truth and significance (p-value) amongst acquisition devices based on Tukey's multiple comparison test

Devices	Precision (%)	Recall (%)	Accuracy (%)	Dice (%)	IoU (%)	95% HD (mm)
CBCT1	0.97±0.02	0.98±0.01	1.00±0.00	0.98±0.01	0.96±0.02	0.82±0.41
CBCT2	0.98±0.01	0.94±0.03	1.00±0.00	0.96±0.02	0.93±0.04	0.90±0.43
CT	0.97±0.01	0.95±0.05	1.00±0.00	0.96±0.03	0.92±0.03	1.79±1.58
Average	0.97±0.01	0.96±0.03	1.00±0.00	0.97±0.02	0.93±0.03	0.98±0.74
p-value						
(CBCT1) vs (CBCT2)	0.3621	0.0011*	1	0.0391*	0.0347*	0.9510
(CBCT1) vs (CT)	0.9902	0.0251*	1	0.0814	0.1063	0.0211*
(CBCT2) vs (CT)	0.6721	0.9578	1	0.9737	0.9978	0.0603*

CBCT 1: Newtom VGi evo, CBCT2: Promax 3D Max, CT: Multi-slice spiral CT, HD:

Hausdorff distance, IOU: Intersection over union

**indicates statistical significance ($p < 0.05$)*