

# ENCODING UPPER NASAL AIRWAY STRUCTURE WITH U-NET FOR RESPIRATORY HEALTHCARE APPLICATIONS

**Bruno A. Pazos, Pablo Navarro, Soledad De Azevedo**

Instituto Patagónico de Ciencias Sociales y Humanas

CENPAT-CONICET

Puerto Madryn, Argentina

{bpazos, pnavarro, deazevedo}@cenpat-conicet.gob.ar

**Claudio A. Delrieux**

Laboratorio de Ciencias de las Imágenes

Universidad Nacional del Sur

Bahía Blanca, Argentina

cad@uns.edu.ar

**Rolando González-José**

Instituto Patagónico de Ciencias Sociales y Humanas

CENPAT-CONICET

Puerto Madryn, Argentina

rolando@cenpat-conicet.gob.ar

## ABSTRACT

The human upper nasal airway is an anatomical structure with a complex geometry that performs essential functions required by the rest of the respiratory system. An accurate and precise segmentation process that captures its intricate shape and variability becomes indispensable to fully understand its performance under different circumstances and to study its anatomy from multiple perspectives. As currently performed, the manual or semi-automatic segmentation process for these structures is extremely time-consuming, may demand extensive manual post-processing steps to correct over- or under-segmentation, and is subject to considerable intra- and inter-operator variance. Further, in developing countries, healthcare institutions modernize their medical imaging devices at different rates; thus, specialists and proposed solutions have to deal with a wide range of image characteristics and quality variability to execute their diagnostics. In this paper we develop an automatic segmentation strategy for the human upper nasal airway, based on a deep convolutional network trained with >3000 CT scans acquired from different devices of a national hospital in Argentina, Hospital Italiano de Buenos Aires (2010). This process achieves remarkable preliminary results with a low error rate (0.07%) and an acceptable similarity score (86.9%).

## 1 INTRODUCTION

One of the main anatomical components of the respiratory system is the upper nasal airway. It performs essential tasks such as conditioning the air before reaching the lungs by regulating its humidity and temperature, or filtering out dust and other potentially dangerous particles. Because of this, it has a complex morphology and presents a wide variability across different ethnic groups and individuals (Huang et al., 2019). In different contexts (clinical or scientific), it is quite useful to analyze geometric representations of this structure (2D or 3D) to understand its functioning process and how it would behave under specific circumstances. Current approaches can perform this task only semi-automatically, requiring operators with wide experience adjusting the initialization parameters and

structures. In addition, after this semi-assisted segmentation is performed, an even more complex post-processing step is required to remove misclassified regions that arise due to limitations and mistakes inherent to the segmentation procedure. This represents a costly and time-consuming process, and can be a daunting task when scans of several individuals must be processed. In particular, the usual feature engineering methods are error-prone due to the complex and sometimes unpredictable shape of this structure. Moreover, several sections of the human skull that participate in the image acquisition present similar intensity levels as the target structure, both in CT and MRI imaging, which gives rise to hard-to-circumvent difficulties in border detection (Kabaliuk et al., 2017; Huang et al., 2016). As a result, traditional segmentation processes may create masks that present considerable amounts of under- or over-segmentation, misrepresenting the real shape of the anatomical structure. In addition, an adequate geometric representation may provide stable features that allow population-wise studies and other research activities that require individual-specific models (Heuzé, 2019; Cellina et al., 2020; Nejati et al., 2016; Deng et al., 2018a). Particularly in Latin America, population-wise studies would allow further understanding of the role of environmental and genetic factors in relation to different health conditions (Dopazo et al., 2019).

On the other hand, recent advances in Deep Learning (DL) yield a variety of neural network architectures for similar segmentation and recognition purposes in 2D and 3D image processing (Badri-narayanan et al., 2017; He et al., 2017). With appropriate training and validation procedures, DL was successfully applied in diverse medical imaging contexts, being able to adequately perform complex segmentation tasks, identify elaborate targets, and detect anomalies in different kinds of anatomical structures and tissues (Ait Skourt et al., 2018; Dong et al., 2011) among other tasks. In this work, we propose the use of a modified version of a convolutional neural network (CNN) that automates the process of segmentation of the upper nasal airway. The presented approach aims to automatically create a 2D representation of the anatomical structure without requiring prior experience or manual assistance, providing reproducible results with minimal error.

## 2 RELATED WORK

Many techniques and methodologies have been proposed recently to address segmentation and geometric reconstruction of the human nasal cavity. Cercos-Pita et al. (2018) presented a tool that combines many of the traditional image processing techniques to achieve a complete 3D representation of the anatomic structure. In a multi-step process, the proposed tool is able to filter metal noise, segment each slice, subtract undesired portions of pixels/voxels that do not belong to the cavity, and lastly, close the holes in the 3D model to create a watertight representation of the structure. Bui et al. (2015) proposed a multi-step level-set framework to automatically segment nasal cavity and paranasal sinuses, creating 3D models of the structure. The process involves the use of thresholds and mathematical morphology operators to locate the RoI, a global active contour process guided by a level-set framework to segment the structure, and localized active contour processes to refine the results. Similarly, Neelapu et al. (2017) developed an algorithm to extract the VoI by automatically identifying landmarks in the structure and initializing the active contour segmentation process. The proposed algorithm presented problems when boundary limits for segmentation regions were not precisely defined. Other authors (Dastidar et al., 1999; Seo et al., 2010; Deng et al., 2018b) have presented semi-automatic approaches to create segmentations for the nasal cavity combining traditional methods. Although the above-mentioned works may achieve good results in some cases, they have limitations and drawbacks due to their semi-supervised nature. Some of them require expert intervention to set and adjust the initial parameters, which, as shown in Cercos-Pita et al. (2018), yields biased results when the human operator is not very experienced.

On the other hand, multiple neural network architectures have been designed in order to create segmentations for different kinds of tissues, tumors and lesions (Ait Skourt et al., 2018; Dong et al., 2011; Suzuki, 2017; Harrigan et al., 2014). For the upper nasal cavity in particular, Xie et al. (2021) have developed a similar approach to the proposed in this work, obtaining 2D as well as 3D representations of the structure, although it was designed to work with MRI instead of CT Scans. Qin et al. (2019) created an architecture that exploits image features in CT Scans to predict airway connectivity by learning the relationship between voxels creating a 3D model of the lower region of the airway. However, DL-based methods tend to perform poorly when the models are tested with datasets that lack diversity. Training with images acquired with different devices will allow the algorithms to deal with image acquisition variance, such as uneven quality, different resolution,

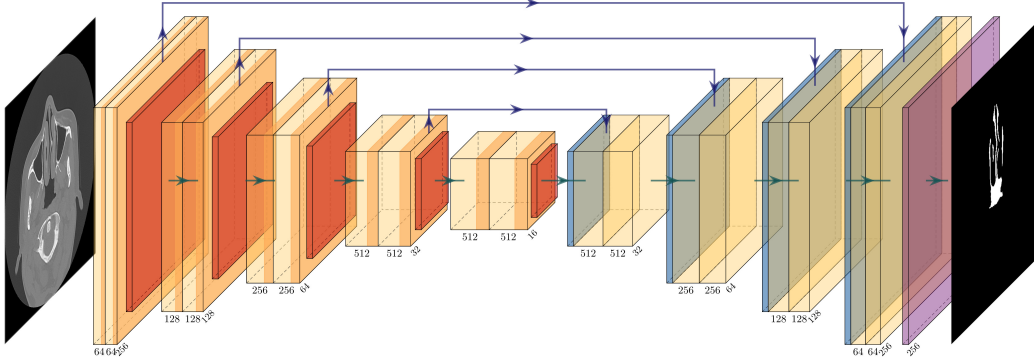


Figure 1: U-Net architecture. The descending path (left) contains contracting operations while the ascending path (right) expands the obtained features to create the segmented mask. Missing information is gathered from skip connections (grey arrows) to achieve better results.

and varying noise levels. Our work proposes an automatic DL method trained and tested with a population-specific dataset (Argentina) with images captured with different devices with ranging image quality input data. The proposed approach performs accurately under the mentioned scenarios while maintaining a low error rate.

### 3 DEEP LEARNING BASED SEGMENTATION MODEL

We propose an automatic approach to segment the anatomical structure, providing a cost-effective solution to the problems mentioned without requiring great effort to use. For this, we applied the U-Net (Ronneberger et al.) architecture, which was successful in other segmentation tasks on medical images. The U-Net model is characterized by three sections: a *descending path*, the encoded data and an *ascending path* (Fig. 1). The dataset is composed by a total of 3100 slices of CT and CBCT scans of human skull in DICOM format from a group of 100 people (64 female, 36 male from ages between 19 and 97) from a national healthcare institution in Buenos Aires. These scans were acquired with different devices under their own individual settings providing images with a diverse set of attributes as mentioned in previous sections of this work. Every CT sequence was further analyzed to only keep the slices in which the nasal cavity was present. The remaining slices were not used nor manually segmented for the training process. The dataset also includes a binary mask for every slice, created manually under the supervision of an expert. Each mask contains black pixels to represent the background and white pixels for the area of interest. Both the dataset and the architecture will be made available by the time of publication.

#### 3.1 IMPLEMENTATION DETAILS

We used Python 3.8 <sup>1</sup> and the Pytorch 1.7 <sup>2</sup> library to clean and prepare the dataset as well as to implement and train the neural network. As the optimizer for our model, we used Adam Kingma & Ba (2015) with the following hyperparameters: *learning\_rate* =  $1e - 5$ , *weight\_decay* =  $1e - 3$ , *beta1* = 0.6 and *beta2* = 0.999. Each training batch was composed of 15 images DICOM and their correspondent binary mask. The U-Net model was trained 500 epochs for approximately 6 hours on a PC with a Nvidia GTX 1070 GPU, an Intel i7-7700 CPU and 16GB of RAM. To quantify the segmentation error of our neural network, we used Binary Cross Entropy (BCE) as our loss function given the binary nature of our expected outputs.

$$L = -\frac{1}{n} \sum_{i=1}^n y_i * \log \hat{y}_i + (1 - y_i) * \log(1 - \hat{y}_i) \tag{1}$$

<sup>1</sup><https://www.python.org/>

<sup>2</sup><https://pytorch.org/>

Table 1: Test Results

Metric	DeepLabv3 (ResNet50)	DeepLabv3 (ResNet101)	Our Method
BCE	0.034 (+/- 0.013)	0.033 (+/- 0.012)	0.007 (+/- 0.001)
DSC	0.659 (+/- 0.085)	0.656 (+/- 0.079)	0.869 (+/- 0.063)
Precision	0.244 (+/- 0.132)	0.242 (+/- 0.133)	0.715 (+/- 0.102)
Recall (Sensitivity)	0.982 (+/- 0.046)	0.981 (+/- 0.050)	0.936 (+/- 0.073)
Specificity	0.983 (+/- 0.006)	0.983 (+/- 0.006)	0.996 (+/- 0.001)
False positive rate	0.016 (+/- 0.006)	0.016 (+/- 0.006)	0.003 (+/- 0.001)

Where  $L$  is the error or loss,  $\hat{y}_i$  is the  $i$ -th value in the model output,  $y_i$  is the corresponding target pixel value (0 or 1), and  $n$  is the output vector size. Optimal values for  $L$  are close to 0.

To measure how close the predicted masks are against the ground truth, we applied Dice Coefficient (DSC)

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}, \quad (2)$$

where DSC is the Dice coefficient for the inputs,  $X$  and  $Y$  represent the predicted output mask and the expected mask accordingly. During implementation, we added an extra term to provide numerical stability. Optimal values for DSC are close to 1.

We scaled down the size of the images to  $256 \times 256$  px (half of its original size) and normalized their values to be between 0 and 1. We also applied *data augmentation* to extend the size of our dataset by introducing random transformations to the input images during training. The initial dataset was divided in 3 subsets with the following proportions: training 70%, validation 10% and testing 20%. The training set was used to modify the weights and parameters of the model, using the validation set after each batch of training images to ensure the learning process its executing correctly or use it as an early-stopping point. Finally, the testing set was used to evaluate the trained model against images never seen during training and to calculate the performance metrics.

## 4 RESULTS

During the experimental phase, we used the testing set of data composed by 219 images. For each image, we calculated both Binary Cross Entropy Loss (BCE) and Dice Coefficient (DSC). Additionally, we also measured Precision, Recall, Specificity and False positive rate. To compare results we trained and evaluated another Deep Learning-based model designed for semantic segmentation purposes, DeepLabv3. We analyzed the performance on two of its variants: DeepLabv3 with ResNet50 backbone and DeepLabv3 with ResNet101 backbone (See Table 1). Although the results in Fig. 2 demonstrate that DeepLabv3 predictions are close to the overall shape of the structure with both variants, it can be noted that several critical features are missing from the segmentation. In contrast, our method achieves segmentation masks with high level of detail, capturing in most cases an almost exact shape of the structure when compared to manually segmented masks.

## 5 DISCUSSION AND CONCLUSION

Our proposal achieves high similarity scores when compared with manually created segmentations, and maintains a low error rate. However, the model may not perform as expected when a complete sequence of CT slices are given as input. This is due to the fact that to avoid the imbalance problem we intentionally left out slices that did not contained the region of interest. As a result, the neural network was not trained to learn to select which slices contain or not portions of the region of interest as well as to filter metal noise or other potential image perturbations. This problem can be easily solved in combination with other algorithms or networks, classifying which slices need to be segmented beforehand as well as filtering noise or removing perturbations. Another limitation is that the model presents under-segmentation results when the portions of region of interest in a slice are too small.

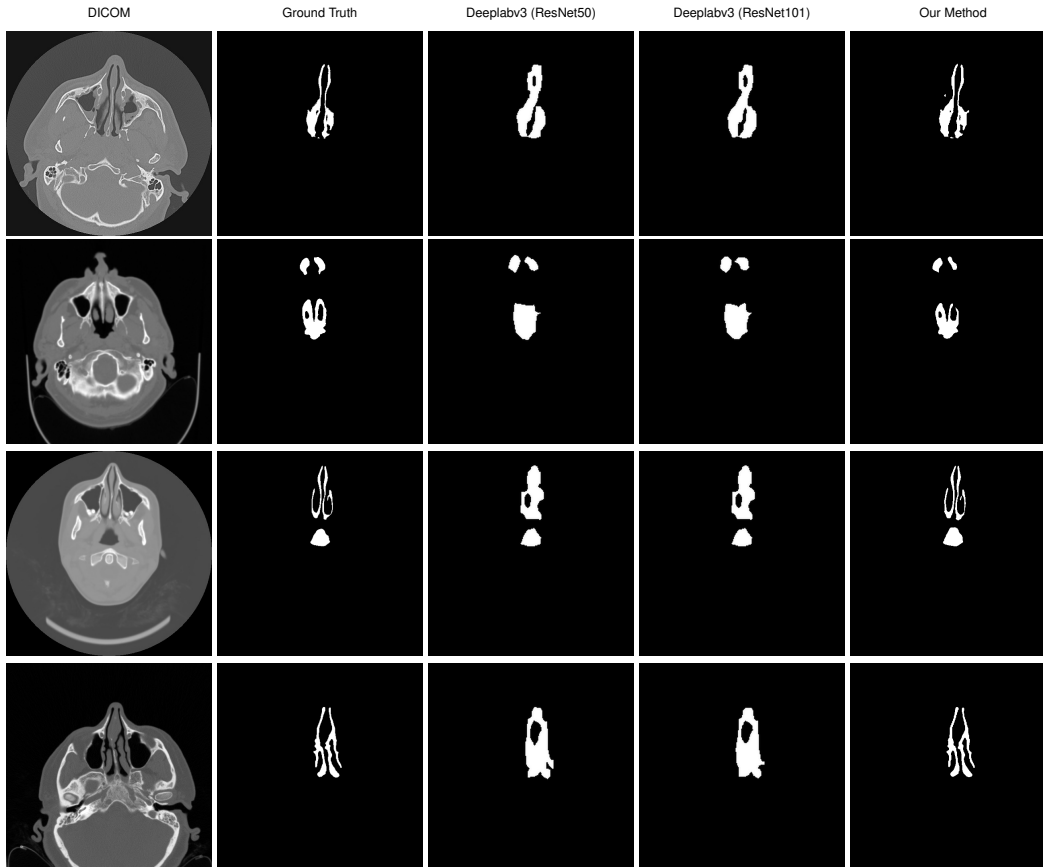


Figure 2: From left to right: DICOM original slices, Ground truth masks, DeepLabv3 (ResNet50) prediction masks, DeepLabv3 (ResNet101) prediction masks and Our method’s prediction masks.

Although the presented results may be valid for 2D analysis, we plan to work in a 3D implementation to complete the framework and for it to be ready to use in the real world. This approach can be the basis of novel computational tools aimed to improve several processes related to surgical procedures that can be simulated *in silico* prior to the intervention itself among other scenarios in daily medical practice, optimizing planning time and reducing the overall cost of such procedures.

Finally, we aim to implement the mentioned algorithms in the near future integrating them with the presented architecture in a framework that automates the whole process.

## REFERENCES

- Brahim Ait Skourt, Abdelhamid El Hassani, and Aicha Majda. Lung CT image segmentation using deep neural networks. In *Procedia Computer Science*, volume 127, pp. 109–113. Elsevier B.V., jan 2018. doi: 10.1016/j.procs.2018.01.104.
- Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2017. ISSN 01628828. doi: 10.1109/TPAMI.2016.2644615.
- Nhat Linh Bui, Sim Heng Ong, and Kelvin Weng Chiong Foong. Automatic segmentation of the nasal cavity and paranasal sinuses from cone-beam CT images. *International Journal of Computer Assisted Radiology and Surgery*, 10(8):1269–1277, aug 2015. ISSN 18616429. doi: 10.1007/s11548-014-1134-5.
- Michaela Cellina, Daniele Gibelli, Annalisa Cappella, Carlo Martinenghi, Elena Belloni, and Giancarlo Oliva. Nasal cavities and the nasal septum: Anatomical variants and assessment of features with computed tomography. *Neuroradiology Journal*, 2020. ISSN 19714009. doi: 10.1177/1971400920913763.
- J.L. Cercos-Pita, I.R. Cal, D. Duque, and G. Sanjuán de Moreta. NASAL-Geom, a free upper respiratory tract 3D model reconstruction software. *Computer Physics Communications*, 223: 55–68, feb 2018. ISSN 0010-4655. doi: 10.1016/J.CPC.2017.10.008.
- P. Dastidar, T. Heinonen, J. Numminen, M. Rautiainen, and E. Laasonen. Semi-automatic segmentation of computed tomographic images in volumetric estimation of nasal airway. *European Archives of Oto-Rhino-Laryngology*, 256(4):192–198, 1999. ISSN 09374477. doi: 10.1007/s004050050138.
- Zhuofu Deng, Takahiko Kitamura, Naoki Matsushiro, Hiroshi Nishimura, Zhiliang Zhu, Min Xu, Kun Xiong, and Yen-Wei Chen. Semi-automatic Segmentation of Paranasal Sinuses from CT Images Using Active Contour with Group Similarity Constraints. In Yen-Wei Chen, Satoshi Tanaka, Robert J Howlett, and Lakhmi C Jain (eds.), *Innovation in Medicine and Healthcare 2017*, pp. 89–98, Cham, jun 2018a. Springer International Publishing. ISBN 978-3-319-59397-5. doi: 10.1007/978-3-319-59397-5\_10.
- Zhuofu Deng, Takahiko Kitamura, Naoki Matsushiro, Hiroshi Nishimura, Zhiliang Zhu, Min Xu, Kun Xiong, and Yen-Wei Chen. Semi-automatic Segmentation of Paranasal Sinuses from CT Images Using Active Contour with Group Similarity Constraints. In Yen-Wei Chen, Satoshi Tanaka, Robert J Howlett, and Lakhmi C Jain (eds.), *Innovation in Medicine and Healthcare 2017*, pp. 89–98, Cham, 2018b. Springer International Publishing. ISBN 978-3-319-59397-5.
- Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. Technical report, 2011.
- Hernán Dopazo, Andrea Llera, Mariana Berenstein, and Rolando González José. Genomas, enfermedades y medicina de precisión: un proyecto nacional. *Ciencia, Tecnología y Política*, 2019.
- Robert L. Harrigan, Swetasudha Panda, Andrew J. Asman, Katrina M. Nelson, Shikha Chaganti, Michael P. DeLisi, Benjamin C. W. Yvernault, Seth A. Smith, Robert L. Galloway, Louise A. Mawn, and Bennett A. Landman. Robust optic nerve segmentation on clinically acquired computed tomography. *Journal of Medical Imaging*, 1(3):034006, dec 2014. ISSN 2329-4302. doi: 10.1117/1.jmi.1.3.034006.
- Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-Octob, pp. 2980–2988. Institute of Electrical and Electronics Engineers Inc., dec 2017. ISBN 9781538610329. doi: 10.1109/ICCV.2017.322.
- Y. Heuzé. What does nasal cavity size tell us about functional nasal airways? *Bulletins et Memoires de la Societe d’Anthropologie de Paris*, 31(1-2):69–76, 2019. ISSN 17775469. doi: 10.3166/bmsap-2018-0011.

- Robin Huang, Ang Li, Lei Bi, Changyang Li, Paul Young, Gregory King, David Dagan Feng, and Jinman Kim. A locally constrained statistical shape model for robust nasal cavity segmentation in computed tomography. In *Proceedings - International Symposium on Biomedical Imaging*, volume 2016-June, pp. 1334–1337. IEEE Computer Society, jun 2016. ISBN 9781479923502. doi: 10.1109/ISBI.2016.7493513.
- Robin Huang, Anthony Nedanoski, David F. Fletcher, Narinder Singh, Jerome Schmid, Paul M. Young, Nicholas Stow, Lei Bi, Daniela Traini, Eugene Wong, Craig L. Phillips, Ronald R. Grunstein, and Jinman Kim. An automated segmentation framework for nasal computational fluid dynamics analysis in computed tomography. *Computers in Biology and Medicine*, 115, dec 2019. ISSN 18790534. doi: 10.1016/j.compbiomed.2019.103505.
- N. Kabaliuk, A. Nejati, C. Loch, D. Schwass, J. E. Cater, and M. C. Jermy. Strategies for Segmenting the Upper Airway in Cone-Beam Computed Tomography (CBCT) Data. *Open Journal of Medical Imaging*, 07(04):196–219, sep 2017. ISSN 2164-2788. doi: 10.4236/ojmi.2017.74019.
- Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, dec 2015. URL <https://arxiv.org/abs/1412.6980v9>.
- Bala Chakravarthy Neelapu, Om Prakash Kharbanda, Viren Sardana, Abhishek Gupta, Srikanth Vasamsetti, Rajiv Balachandran, Shailendra Singh Rana, and Harish Kumar Sardana. A pilot study for segmentation of pharyngeal and sino-nasal airway subregions by automatic contour initialization. *International Journal of Computer Assisted Radiology and Surgery*, 12(11):1877–1893, nov 2017. ISSN 1861-6410. doi: 10.1007/s11548-017-1650-1.
- Alireza Nejati, Natalia Kabaliuk, Mark C. Jermy, and John E. Cater. A deformable template method for describing and averaging the anatomical variation of the human nasal cavity. *BMC Medical Imaging*, 16(1):1–12, oct 2016. ISSN 14712342. doi: 10.1186/s12880-016-0154-8.
- Yulei Qin, Mingjian Chen, Hao Zheng, Yun Gu, Mali Shen, Jie Yang, Xiaolin Huang, Yue-Min Zhu, and Guang-Zhong Yang. AirwayNet: A Voxel-Connectivity Aware Approach for Accurate Airway Segmentation Using Convolutional Neural Networks. Technical report, 2019.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. Technical report. URL <http://lmb.informatik.uni-freiburg.de/>.
- Anna Seo, S. K. Chung, Jun Lee, Jee In Kim, and Hyung Seok Kim. Semiautomatic segmentation of nasal airway based on collaborative environment. In *Proceedings - 2010 International Symposium on Ubiquitous Virtual Reality, ISUVR 2010*, pp. 56–59. IEEE, jul 2010. ISBN 9780769541242. doi: 10.1109/ISUVR.2010.24.
- Kenji Suzuki. Overview of deep learning in medical imaging, sep 2017. ISSN 18650341.
- Lipeng Xie, Jayaram K. Udupa, Yubing Tong, Drew A. Torigian, Grace Huang, Rachel M. Kogan, Jennifer Ben Nathan, David M. Wootton, Kokren R. Choy, Sanghun Sin, Mark Wagshul, and Raanan Arens. Automatic upper airway segmentation in static and dynamic MRI via deep convolutional neural networks. In Barjor S. Gimi and Andrzej Krol (eds.), *Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 11600, pp. 16. SPIE, feb 2021. ISBN 9781510640290. doi: 10.1117/12.2581974.