Pulmonary airway segmentation using an 3D anisotropic U-Net ensemble and automated connectivity analysis

Felix Thielke[1]

Fraunhofer Institute for Digital Medicine MEVIS, Max-von-Laue-Str. 2, 28359 Bremen, Germany felix.thielke@mevis.fraunhofer.de

1 Introduction

This work describes a proposed deep neural network (DNN) based solution to the task of segmenting pulmonary airways in the context of the ATM22 challenge. To achieve this task, the proposed approach builds heavily on prior work on DNN based segmentation of hepatic vasculature.

2 Methods

The provided challenge data training set was split into internal training, validation and test sets consisting of 210, 30 and 60 cases respectively. While the validation set was used for monitoring the training progress and model selection, the test set allowed to measure the performance of the resulting algorithm based on several evaluation metrics before submitting it to the leaderboards.

2.1 DNN Training

As architecture for the trained DNN models, the anisotropic 3D U-Net architecture by Chlebus et al. [1] was chosen, since it has proven to perform well for segmentation of vascular structures [5].

To assure that the input data of the trained model consists of the same distribution at both training and test time, the input images are clamped to a fixed minimum value of -1000 HU and resampled such that each voxel has the size $(0.8\text{ mm})^3$.

Similar to Thielke et al. [5], the proposed approach extracts patches of size $52 \times 52 \times 52$ with a padding of $92 \times 92 \times 20$ at each side from the preprocessed volumes so that training and inference can use the overlapping tile strategy as proposed by Ronneberger et al. [3]. To stabilize the training, patches that only contain background voxels are excluded.

The training scheme mostly resembles that of the nnU-Net [2]. Models are trained for 250 000 iterations with two patches per training batch and the loss function is a combination of the categorical crossentropy and soft Dice losses. Unlike the nnU-Net, the learning rate is linearly decayed during the training,

starting at 0.001 and ending at 0.00001 at the final iteration. NovoGrad is used as optimizer. To regularize the training, a feature dropout layer with rate 0.25 is inserted before each convolutional layer in the decoder part of the architecture.

Since the goal is achieving a high segmentation accuracy around the airways, but the performance far from the airway tree is not important—simple postprocessing steps can be used to find the trachea and reconstruct the bronchi tree from there—, a loss weighting scheme was devised that depends on the distance d_v of each voxel to the surface of the foreground mask as shown in equation 1. Weighting based on the distance from the surface additionally has the benefit that it reduces the relative weight of large branches in the loss function, which would otherwise be overrepresented since they consist of many voxels. Therefore this weighting scheme also encourages the models to focus more on branches with smaller diameters.

$$\text{weight}_v = \begin{cases} 1 & \text{if} d_v \leq 3\text{mm} \\ \max\left(0.01, 1 - \frac{d_v}{15\text{mm}}\right) & \text{otherwise} \end{cases} \tag{1}$$

As mentioned above, the validation set is used to select the best model during training: every 1000 training iterations a validation is performed. In the end, the model with the best Jaccard index on the validation set is kept.

To improve generalization, some data augmentation steps were added. Before patch extraction, volumes are randomly rotated around the Z axis by either $0°$, $10°$ or $-10°$ and around each of the X and Y axes by either $0°$ or $2°$. Additionally, gaussian noise is added to each training patch, where $\sigma \in [0, 20]$ is randomly chosen per patch.

After one was trained with and one without data augmentation, averaging the outputs of both models turned out to lead to better segmentation performance than either of the models alone. Thus, in the challenge submission both models are used as an ensemble.

2.2 Postprocessing

To extract a connected airway graph from the raw model, a method originally developed for hepatic vessel graph extraction is used [5], which connects the components of the skeletized output mask using a shortest path algorithm, where the raw softmax outputs of the model are used as the cost image. Unlike the approach for liver vessel segmentation, the method of finding the root point of the graph was modified for pulmonary airway segmentation. To detect the root point—which should be located in the trachea—, first the skeleton of the largest connected component of the model output when binarized with a low threshold of 0.1 is calculated. The root then is the point on the skeleton with the largest associated radius.

2.3 Evaluation metrics

For the validation phase, the ATM22 challenge publishes the Dice score, number of false positives (FP) and number of false negatives (FN). Therefore, this work also uses the Dice score for evaluation.

For truly judging the correctness of the segmentation of pulmonary airways, topological measures are however equally if not more important than volumetric measures. The topological measure used here is the centerline Dice (clDice) [4] as well as its components topology precision (called clPrecision in the following) and topology sensitivity (clSensitivity).

3 Results

Figure 1 shows the results of evaluating the resulting model ensemble on the test set. As can be seen, the median of both Dice and clDice is greater than 0.8. The clSensitivity even has a median of 0.9496, which means that the clDice is only limited by the clPrecision (0.7642). In other words, the model rather predicts too many segments of the graph than too few.

A visual inspection of these additional branches by experts is needed to assess if the method really predicts too much or if the problem lies in the reference data which would then not contain all visible parts of the airways. Figure 2 shows the extracted airway graphs for two cases of the test set.

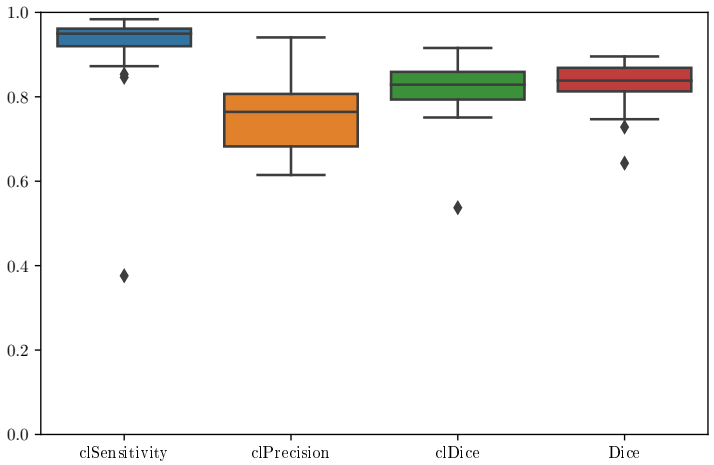

Fig. 1. Metrics for the resulting model on the test set.

Fig. 2. Extracted airway graphs for two example cases from the test set. The algorithm outputs are on the left, the reference masks on the right side.

References

1. Chlebus, G., Schenk, A., Hahn, H.K., Van Ginneken, B., Meine, H.: Robust Segmentation Models Using an Uncertainty Slice Sampling-Based Annotation Workflow. IEEE Access **10**, 4728–4738 (2022). https://doi.org/10.1109/ACCESS.2022.3141021
2. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z
3. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. pp. 234–241. Lecture Notes in Computer Science, Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
4. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P.W., Bauer, U., Menze, B.H.: clDice - A Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. pp. 16560–16569 (2021)
5. Thielke, F., Kock, F., Hänsch, A., Georgii, J., Abolmaali, N., Endo, I., Meine, H., Schenk, A.: Improving deep learning based liver vessel segmentation using automated connectivity analysis. In: Medical Imaging 2022: Image Processing. vol. 12032, pp. 886–892. SPIE (Apr 2022). https://doi.org/10.1117/12.2612526