# ATM 2022 Challenge Submission by cvhthreedee

Constantin Seibold[1], Alexander Jaus[1], Simon Reiß[1], Zdravko Marinov[1],
Matthias Fink[2], and Rainer Stiefelhagen[1]

[1]Karlsruhe Institute of Technology, Vincenz-Priessnitz-Str. 3, 76137 Karlsruhe, Germany
[2]University Clinic Heidelberg, Im Neuenheimer Feld 420, 69120 Heidelberg, Germany

## 1 Proposed Approach

### 1.1 Potential Architectures for the ATM Challenge

There have been a number of works targeting medical image segmentation. One of the earliest works which is still among the most popular baseline models is the UNet [1] which is based on an encoder-decoder architecture. A major contribution of the UNet is to provide the decoder which starts the reconstruction phase with highly semantic aggregated information with the more high level pixel information of the encoder of the matching stage. This leads to the U-shape architecture and has proven to be a very effective design choice especially in the medical domain where the UNet is still the de-facto standard for segmentation task whereas in other task, more sophisticated models such as models of the DeepLab family DeepLab [2, 3].

A major limitation of the UNet is its limited applicability to naturally address 3D images which are common in the medical domain. CT, MRI or PT are among the most widely used imaging methods which all render images in 3 dimensions. This limitation of the UNet is addressed in the 3D-Unet [4] paper and the V-Net [5] paper which extend the Unet to 3D images.

Unet++ [6] which aims to bridge the gap between semantically dissimilar features which are concatenated in the standard Unet by introducing intermediate convolutional blocks between the encoder and the decoder.

Lately many works have focused on bringing transformers [7] after their successful application in the area of computer vision [8–11] to the medical image domain by introducing UNETR [12] and Swin UNETR [13].

### 1.2 Efficiency Constraints

We strongly believe in the power of neural network architectures which directly on 3D images, as slicing cubes into planes always provides a limited view of the given possibilities by the 3D modality and in theory information fusion of different angles in order provide the network with the structural information. A major drawback of 3D based neural networks is their larger requirements for training and inference time. This is easily imaginable as the typical size of a 3D $H \times W \times D$ image contains $D$ 2D images, where $D$ is often multiple hundred slices.

Due to the hardware requirements during inference time restricted to one NVIDIA GEFORCE RTX 2080ti we refrained from the usage of a 3D model despite its superior performance shown during our internal cross validation and aimed to maximize the performance of a 2D model. We furthermore support the usage of the more standard 2D Unet due to the efficiency incentives posed by the hosts. Due to the inference time constraints we refrain from model ensembling and respective postprocessing.

### 1.3 Chosen architecture

As the nnUnet [14] work has shown, for many applications in the medical domain the correct preprocesssing and the choice of hyper parameters is often times even more important than a adapted model architecture. We follow this thought and chose to work with a standard 2D Unet which matches the posed hardware restrictions.

We follow the proposed approach of the nnUnet and end up with a 2D Unet. We train for 1000 epochs using the SGD optimizer and a learning rate of $1e-3$ using a dice and cross entropy loss. Patches seen during training are ensured to contain a positive label. We apply an extraction of the largest non-background connected component to comply with the challenge evaluation setup.

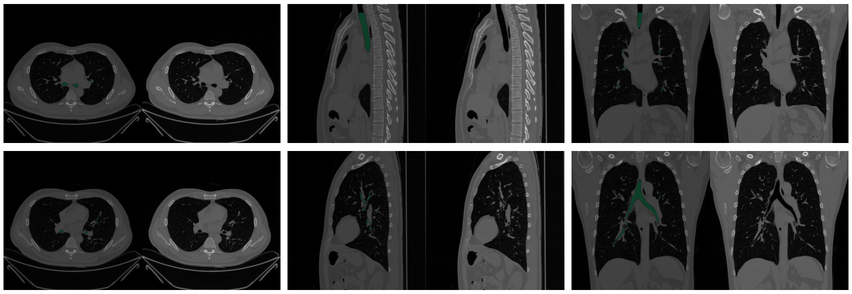

**Fig. 1.** Qualitative results with prediction (left) and input (right) of the 2D UNet on ID293 of our validation set in axial, sagittal, and coronal views.

## 2 Preliminary Results

We display some qualitative results on multiple views in Fig. 1. We see that proper segmentations on larger connected regions such as the trachea. We noticed most errors stem from the model at times fails to pick up thin areas in the airway tree. As such the network predictions lead to multiple connected components of which we choose the largest. To combat this, we experimented with postprocessing such as region growing under certain threshold rules, however, these did not improve performance.

Our chosen UNet achieves a dice score of 0.8603 on a holdout validation set within reasonable time and memory constraints.

## References

1. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention.* Springer, 2015, pp. 234–241.
2. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
3. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
4. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention.* Springer, 2016, pp. 424–432.
5. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV).* IEEE, 2016, pp. 565–571.
6. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support.* Springer, 2018, pp. 3–11.
7. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
8. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
9. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 012–10 022.
10. E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12 077–12 090, 2021.
11. S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. Torr *et al.*, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 6881–6890.
12. A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 574–584.

13. A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI Brainlesion Workshop*. Springer, 2022, pp. 272–284.
14. F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.