Comparing Image Segmentation Neural Networks for the Analysis of Precision Cut Lung Slices

Mohan Xu[1], Susann Dehmel[1], Lena Wiese[1,2],

[1]Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany
[2]Institute of Computer Science, Goethe University Frankfurt, Frankfurt a. M., Germany
mohan.xu@item.fraunhofer.de

Abstract. Bronchodilators serve as a pivotal intervention for ameliorating symptoms associated with Inflammatory and allergic lung diseases. The objective assessment of bronchodilator efficacy is critical for therapeutic optimization. Measuring airflow volume through precision cut lung slices (PCLS) imaging at varying time intervals provides a quantitative means to assess airway patency. To enhance the efficiency of this evaluation process, our study extends the existing image segmentation workflow to encompass a wider range of neural networks. Extensive experiments have been conducted across varied data preprocessing methods and loss functions. Furthermore, we contrast the performance differences between single and ensemble models, alongside a visual comparative analysis of their detailed variances in image segmentation. This refined workflow not only surpasses previous experimental results but also enhances the accuracy of lung treatment programs, offering a broader array of choices for future image segmentation tasks.

1 Introduction

Bronchodilators relax bronchial smooth muscle, dilate the bronchi, and alleviate airflow limitations. They constitute a vital therapeutic tool for managing inflammatory and allergic lung diseases. Evaluating the effectiveness of various bronchodilators can be achieved by measuring the airflow capacity of PCLS. During image acquisition, factors such as petri dish movement, changes in illumination conditions, out-of-focus images, or obscured airways can adversely impact the image quality of PCLS [1], affecting the accuracy of assessment results. Hence, this study focuses on the precise and efficient segmentation and calculation of airway areas in images of varying quality.

Deep learning, as a branch of machine learning, boasts remarkable capabilities in pattern recognition and feature learning, making it highly applicable in medical image analysis. Within the realm of medical image analysis, image segmentation stands as a crucial application, giving rise to a plethora of neural networks that exhibit outstanding performance. The skip-connection structure, introduced in [2, 3], establishes connections between the encoder and decoder, harnessing information at multiple levels. Furthermore, it seamlessly integrates shallow and deep features to bridge potential semantic gaps. Atrous convolution, as discussed in [4, 5], enhances the receptive field of feature map points without introducing additional parameters. PSPNet [6] achieves rapid segmentation results while maintaining high-quality image segmentation. Linknet [7] and FPN [8] share similarities with Unet in their architectures. However, Linknet adopts the resnet structure for its encoder block, while FPN accomplishes feature fusion

by up-sampling feature maps at varying scales. In PAN [9], the FPA module significantly expands the receptive field, while the GAU module extracts global context from high-level features, guiding the weighting operation of low-level features. MAnet [10], similar to the Unet structure, seeks to capture spatial dependencies between pixels on a global scale and channel dependencies among arbitrary feature maps through multi-scale semantic feature fusion.

Utilizing the bronchoconstriction dataset collected from PCLS, [11] employed Unet for image segmentation across four categories: background, airway, blood vessel, and airway boundary. This approach significantly accelerates image segmentation compared to manual methods. [1] expanded the scope of user-oriented web applications to include Unet, PSPNet, LinkNet, and FPN, constructing a comprehensive end-to-end image analysis framework. This study introduces new data preprocessing methods, neural network architectures and loss functions building upon prior research, ultimately yielding superior outcomes across a spectrum of evaluation metrics. Moreover, we conduct a comparison of segmentation performance between single and ensemble models, demonstrating that the ensemble approach enhances the model's ability to discriminate image details without incurring extra training costs.

2 Materials and Methods

2.1 Dataset and experimental setup

Inverted microscopes and digital video cameras were employed to capture images of the airway region in PCLS samples contained in petri dishes, which had been treated with bronchodilators. The experimental design encompassed a wide range of variations in compounds, doses and samples, resulting in a total of 420 different experimental setups. The bronchoconstriction dataset, comprising 420 images, consists of images captured under various experimental conditions. Each image includes components such as background, airway, airway boundary, and possible blood vessels.

In this study, we employed 9 popular image segmentation networks defined in the segmentation models pytorch [12], which include Unet, Unet++, Deeplabv3, Deeplabv3+, FPN, Linknet, PSPNet, PAN, and MAnet. The training process for these models utilized an SGD optimizer, trained for 80 epochs with a learning rate of 0.1. Early Stopping was implemented with a threshold set to 20 epochs. To ensure the comparability of experimental results, we adopted the dataset partitioning approach from references [1, 11] and conducted 10-fold cross-validation.

2.2 Proposed workflow

In this section, we detail the enhanced workflow for image segmentation. Fig. 1 illustrates the process wherein a given input image, along with its corresponding ground truth, is subjected to a preprocessing step before being fed into a neural network for training. Following this, the neural network's output is processed through a softmax function prior to the computation of the loss function against the ground truth. Distinct colors in the output image represent the various components of the PCLS. Furthermore, images

can undergo testing using either a single neural network or ensemble neural networks by adjusting the respective weights.

The preprocessing step encompasses resize, online data augmentation and normalization. To enhance the dataset, which was resized to 512x512, we employed 6 methods from the albumentations library [13] for online data augmentation of the training set, as illustrated in Fig. 1(a-f). These data augmentation methods can be applied independently or in combination to the images, with the application likelihood and intensity modifiable through parameter adjustments. Normalization serves to equalize the variations among disparate data volumes and to optimize the data distribution, thereby accelerating the convergence of the model.

The nine neural networks utilized for the image segmentation task comprise an encoder, a decoder, and a segmentation head. The encoder is charged with extracting both global and local features from the image. We selected the pretrained *resnet101* model as the encoder to expedite model convergence. The decoder incrementally enhances the resolution of the feature maps through up-sampling or deconvolution, integrating them with feature maps from various stages of the encoder. Positioned subsequent to the decoder, the segmentation head guarantees that the output channel count of the network corresponds to the number of categories to be predicted. Images are tested by using trained neural networks with varying weights, adhering to the condition: $\sum_{i=1}^{n} w_i = 1$. When testing is conducted on a single model, the weight assigned to that model is set to 1, while the weights for all other models are set to 0.

The loss function, a vital indicator of model performance in deep learning, assesses the degree of divergence between the model's predictions and the actual values. During

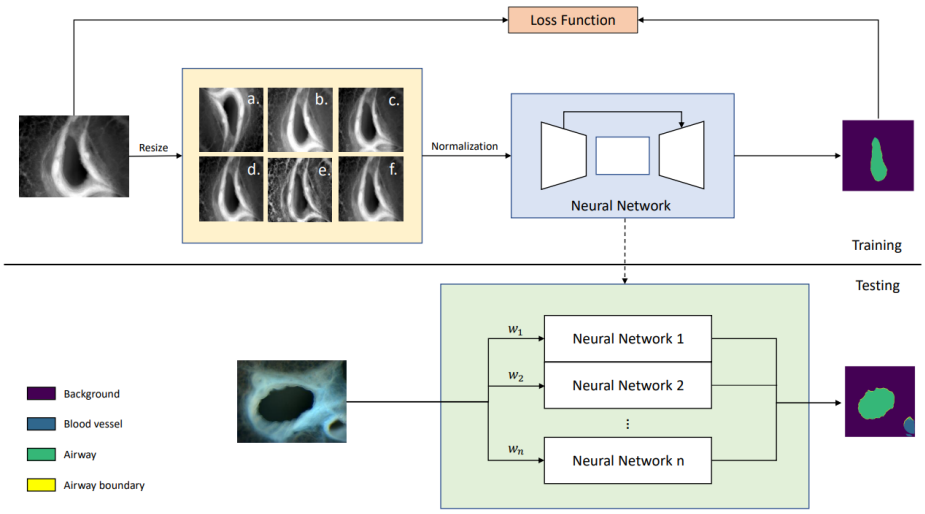


Fig. 1. Image segmentation workflow based on bronchoconstriction dataset, where a-f correspond to 6 data augmentation methods: Flip, BrightnessContrast, GridDistortion, OpticalDistortion, CLAHE and RandomGamma. The annotations in the lower-left indicate the categories associated with the various colored regions in the predictions.

Tab. 1. Comparative analysis of Unet performance with various data augmentation methods.

	Original	a	a+b	a+c	a+d	a+e	a+e+f
IoU	94.95	95.40	95.30	94.87	94.63	95.48	95.55
Dice	97.37	97.61	97.56	97.32	97.19	97.66	97.71
ASD	5.02	3.00	2.82	3.35	3.62	2.95	2.80

Tab. 2. Comparative analysis of Unet performance with various loss functions.

	Cross entropy loss	Focal loss	Dice loss	Tversky loss
IoU	94.95	92.44	94.66	94.79
Dice	97.37	96.01	97.20	97.27
ASD	5.02	5.75	6.89	5.39

the training of neural networks, we examined the model's performance across four loss functions: cross-entropy loss, focal loss, dice loss, and tversky loss, comparing the variances in outcomes to identify the most effective loss function.

3 Results

To determine the optimal data augmentation and loss function combinations, we conducted a series of experiments using Unet. The effects of the 6 data augmentation methods, as identified in Fig. 1, are presented in Tab. 1. Notably, the inclusion of the horizontal and vertical flip methods significantly enhanced the model's performance. The implementation of the Flip method resulted in improvements of 0.45% in IoU and 0.24% in Dice coefficients, along with a decrease of 1.98 in the average surface distance (ASD). Subsequent experiments, building upon the Flip method, determined that a combination of Flip, CLAHE, and Gamma augmentations (a+e+f) yielded the highest performance on the bronchoconstriction dataset. Concerning loss functions, Tab. 2 presents a comparative analysis of their impact on model performance when no data augmentation is applied. The model utilizing cross entropy loss surpassed other loss functions across all metrics. Relative to the least effective loss function, the cross entropy loss registered enhancements of 2.51% in IoU, 1.26% in Dice, and 0.73 in ASD.

Tab. 3 presents the performance outcomes derived from 10-fold cross-validation of nine neural networks, utilizing a consistent random seed. The enhanced workflow demonstrates superior performance over previous work in terms of IoU and Dice coefficients, particularly for the Unet architecture, which exhibits improvements of 1.94% and 1.11%, respectively. Furthermore, additional architectures such as Unet++, MAnet, PAN, Deeplabv3, and Deeplabv3+ were employed for image segmentation tasks within the bronchoconstriction dataset. Of these, Deeplabv3+ achieved the most favorable results across all three evaluation metrics. In the ensemble experiments, we compared three ensemble strategies (Ensemble 1 contains Deeplabv3+ and Unet++; Ensemble 2 contains Deeplabv3+, Unet++, Unet, FPN and Deeplabv3; Ensemble 3 contains Deeplabv3+, Unet++, Unet, FPN, Deeplabv3 and MAnet). Each assigning equal weights to the models. The findings reveal that while the ensemble approach surpasses Deeplabv3+ in

Tab. 3. Comparative analysis of neural network performance: our work versus previous work.

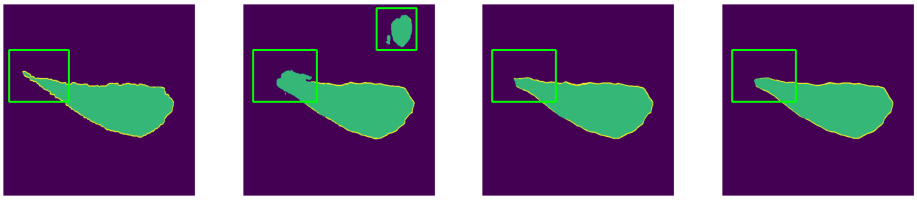
	Previous work [1]		Our work		
	IoU	Dice	IoU	Dice	ASD
Unet	93.14	96.34	95.08±0.32	97.45±0.17	3.74±0.44
PSPNet	92.95	96.27	93.14±0.45	96.38±0.24	4.03±0.55
Linknet	93.43	96.52	94.92±0.40	97.36±0.21	5.09±1.32
FPN	94.16	96.91	95.00±0.26	97.41±0.14	3.44±0.36
Unet++		-	95.14±0.47	97.47±0.25	3.73±0.53
MAnet		-	94.96±0.27	97.38±0.15	3.55±0.37
PAN		-	94.65±0.56	97.22±0.27	3.65±0.78
Deeplabv3		-	95.00±0.28	97.40±0.15	3.63±0.43
Deeplabv3+		-	95.18±0.43	97.50±0.23	3.11±0.45
	Ensemble 1		95.42±0.29	97.63±0.16	3.08±0.29
	Ensemble 2		95.55±0.20	97.70±0.11	2.79±0.13
	Ensemble 3		95.53±0.21	97.69±0.11	2.79±0.16

terms of overall performance, indiscriminately increasing the number of models does not guarantee improved image segmentation results. Fig. 2 illustrates the performance of the single and ensemble models in segmenting background, airway and airway boundary.

4 Discussion

In this study, we executed a series of experiments on the bronchoconstriction dataset, extending the existing workflow across four dimensions: data processing methods, neural network architectures, loss functions, and evaluation metrics. These experiments aimed to identify the optimal training strategy. The results indicate that the enhanced workflow surpasses its predecessors in performance and offers a range of options for improving the performance of image segmentation tasks. Within the ensemble strategy, the assignment of model weights accurately reflects each model's contribution to overall performance. Notably, merely increasing the number of models while reducing the weights of existing models does not positively impact model performance.

Given that images in the bronchoconstriction dataset featuring the blood vessel category comprise merely 7.4% of the dataset, the influence of category imbalance on



(a) Ground truth **(b)** Deeplabv3+ **(c)** Ensemble 1 **(d)** Ensemble 2

Fig. 2. Comparative visualization of image segmentation on single and ensemble models.

model performance cannot be overlooked. In view of this, future work will address this imbalance issue to improve the segmentation of minority categories within the dataset.

Acknowledgement. This work was supported by the Fraunhofer Internal Programs under Grant No. Attract 042-601000.

References

1. Wiese L, Höltje D. NNCompare: a framework for dataset selection, data augmentation and comparison of different neural networks for medical image analysis. Proceedings of the Fifth Workshop on Data Management for End-To-End Machine Learning. 2021:1–7.
2. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer. 2015:234–41.
3. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer. 2018:3–11.
4. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. 2017.
5. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV). 2018:801–18.
6. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:2881–90.
7. Chaurasia A, Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation. 2017 IEEE visual communications and image processing (VCIP). IEEE. 2017:1–4.
8. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:2117–25.
9. Khosravan N, Mortazi A, Wallace M, Bagci U. Pan: Projective adversarial network for medical image segmentation. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. Springer. 2019:68–76.
10. Li R, Zheng S, Zhang C, Duan C, Su J, Wang L et al. Multiattention network for semantic segmentation of fine-resolution remote sensing images. IEEE Transactions on Geoscience and Remote Sensing. 2021;60:1–13.
11. Steinmeyer C, Dehmel S, Theidel D, Braun A, Wiese L. Automating Bronchoconstriction Analysis based on U-Net. EDBT/ICDT Workshops. 2021.
12. Iakubovskii P. Segmentation Models Pytorch. `https://github.com/qubvel/segmentation_models.pytorch`. 2019.
13. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: Fast and Flexible Image Augmentations. Information. 2020;11(2).