Proceedings of Machine Learning Research vol 245:1–10, 2024 2024 International Conference on Machine Learning and Intelligent Computing

Multiscale Fusion and Boundary Refinement UNet Model for Lung Parenchyma Segmentation

He-He Huang
Ocean College, Jiangsu University of Science and Technology, Zhenjiang, 212100, China

Sen-Yu Wei* 15252928679@163.COM
Department of Respiratory Medicine, the Affiliated Hospital of Jiangsu University, Zhenjiang 212001, China

Chen Zhao
Ocean College, Jiangsu University of Science and Technology, Zhenjiang, 212100, China

Yu Shi
School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 610054, China

Yuetao Zhao* ZHAOYUETAO@JUST.EDU.CN
Ocean College, Jiangsu University of Science and Technology, Zhenjiang, 212100, China
School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 610054, China

Yuan Li
Ocean College, Jiangsu University of Science and Technology, Zhenjiang, 212100, China

Qiu-Ming Zang
Ocean College, Jiangsu University of Science and Technology, Zhenjiang, 212100, China

Jianhua Xu
School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 610054, China

Editors: Nianyin Zeng and Ram Bilas Pachori

Abstract

Methods based on automatic computer-aided systems for lung cancer diagnosis have gained popularity in recent years. Lung parenchyma segmentation technology plays an important role in these methods. To reduce the gradient loss, improve the feature utilization, and alleviate the difficulty of fully mining the contextual information in lung parenchyma segmentation, this study proposes a network model for lung parenchyma segmentation based on a Multiscale fusion and Boundary Refinement UNet (MBR–UNet) model. First, along the encoding path, the features are efficiently extracted by a Residual–Residual block module. Next, along the decoding path, a multiscale attention–spatial pyramid pooling module fully integrates the feature maps of different layers and sums the outputs of each layer of the decoding path for boundary refinement. Finally, the training model is optimized through a hybrid loss function. The proposed model is experimentally evaluated on the lung segmentation dataset of the Kaggle competition. The accuracy, Dice similarity coefficient, intersection ratio, and Hausdorff distance of the network segmentation are improved by 98.79%, 97.35%, 96.34%, and 12.82 mm, respectively, from those of other segmentation methods. According to these results, the method can more precisely segment pulmonary parenchyma than the existing methods.

Keywords: multiscale fusion, Boundary refinement, UNet, Lung segmentation

1. Introduction

Early diagnosis and intervention are crucial for improving the clinical outcome of lung cancer, a malignant tumor that seriously endangers human health (Sung et al., 2021). U-shaped codecs and jump connection structures, represented by the UNet model, are widely adopted in lung parenchyma segmentation algorithms (Skourt et al., 2018). Consequently, improving the UNet network has become a major focus of lung-tissue segmentation research. Geng et al. (2019) improved the segmentation capability of the UNet network by combining VGG-16 with inflated convolution. Khanna et al. (2020) enhanced the UNet network with deep residual blocks that extract more discriminative features and data enhancement techniques that improve the generalizability of the network. The UNet-based dual-stream lung-tissue segmentation network of Zhu et al. (2024) introduces shape stream branches and uses the deep features of the backbone network to generate shape streams with clear edge information, thereby improving the segmentation performance. Although the UNet network has undoubtedly contributed to medical image segmentation, it is hampered by gradient loss, low feature utilization, and difficulty in fully mining the contextual information in medical images with unclear boundary features.

To alleviate the problems of traditional UNet networks, this paper proposes a pulmonary parenchymal segmentation network model based on Multiscale fusion and Boundary Refinement UNet (MBR–UNet). The model adopts the U-shaped codec and a similar hopping connection structure to that of UNet but has different core components.

2. Methods

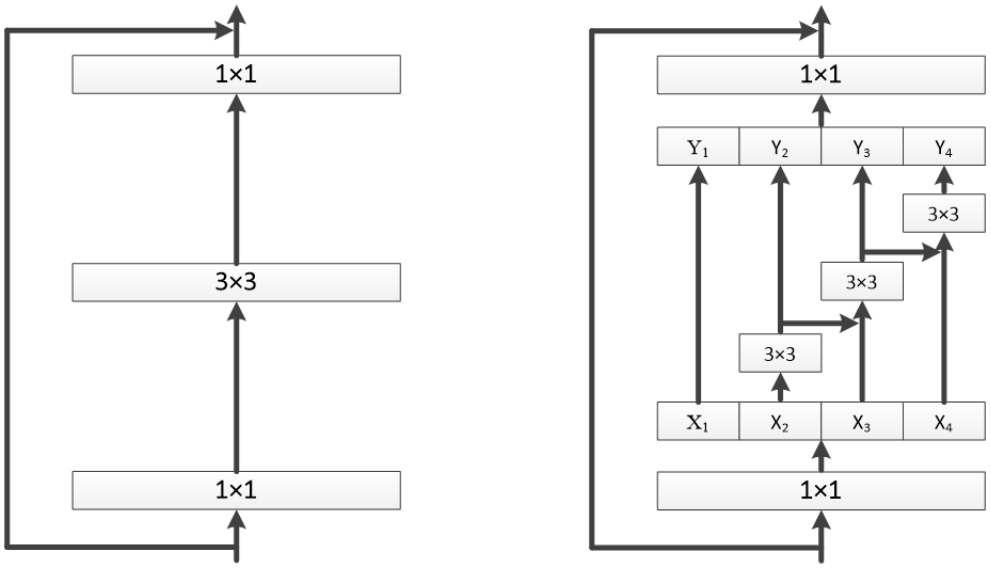
2.1. Res2Net Block Module

The Residual–Residual Network block (Res2Net block) is a new module proposed by Gao et al. (2019). Inspired by the Residual Network (ResNet), Res2Net extracts multiscale features from images by simply modifying the Bottleneck block in ResNet. Specifically, Res2Net block adopts a split–transform–merge structural design (Figure 1(b)) that disperses and processes information in parallel, thus alleviating the bottlenecks in the information transfer process of the traditional ResNet structure. By introducing multi-level feature interactions, Res2Net Block effectively enhances the network's ability to perceive multiscale features, improving its performance in image recognition tasks.

2.2. Multiscale Attention to Spatial Pyramid Pooling Modules

The decoding path of the UNet network fuses shallow and deep feature maps through simple upsampling and jump connections, which hinders the effective integration of features at different levels. Moreover, as the convolutional kernel size and step size of the decoding part are fixed, UNet cannot easily perform multiscale segmentation tasks that must capture global information. Chen et al. (2018) inserted inflated convolutional branches and globally averaged pooling branches into pooling modules that capture the detailed local information and fuse the global contextual information. The Multiscale Convolutional Attention (MSCA) module, proposed by Guo et al. (2022) in 2022, aggregates the local information using deep convolution to improve the image quality. Inspired by the MSCA and pooling modules, this paper designs a Multiscale Attention–Spatial Pooling Pyramid (MSA–SPP) module with the structure shown in Figure 2. First, the MSA–SPP inputs feature

maps into a global pooling branch and three expansion convolution branches with different expansion rates, which capture the global and local semantic information and the contextual information at different scales in the feature maps. The feature maps output from these four branches are then spliced and input to three deep-strip convolution branches with different kernel sizes, which enhance the feature expression ability at a finer granularity level. Finally, the feature maps are extracted with multiple multiscale contextual information and the original input feature maps are spliced with 1×1 convolution for channel adjustment, enabling the complete fusion of high- and low-level feature maps.



(a) Bottleneck block (b) Res2Net block

Figure 1: Comparison of the structures of (a) the Bottleneck block in traditional ResNet and (b) Res2Net block.

2.3. Boundary Refinement Network

A typical refinement module (RM) (Deng et al., 2018) is a residual block that finely segments the residuals between the rough and real graphs; however, the shallow network hierarchy of the existing RMs impedes the extraction of fine, high-level details. To avoid this problem, Qin et al. (2021) in 2019 proposed the residual refinement module (RRM), which aims to optimize the unevenly predicted regions and fuzzy boundary defects in the coarse feature graph. The RRM structure is shown in Figure 3.

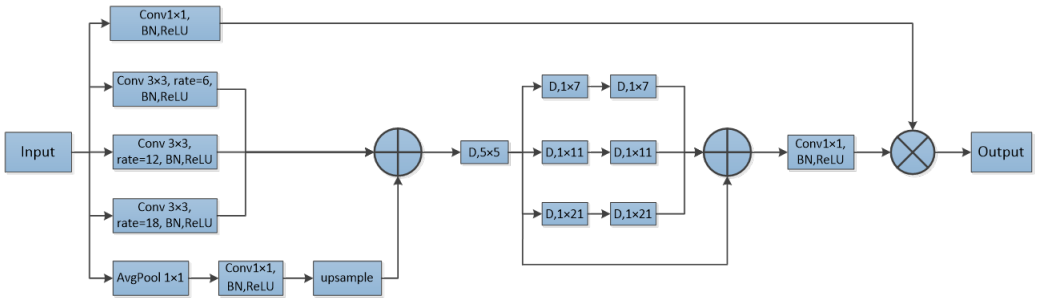


Figure 2: Structure of the proposed Multiscale Attention–Spatial Pooling Pyramid (MSA–SPP) module.

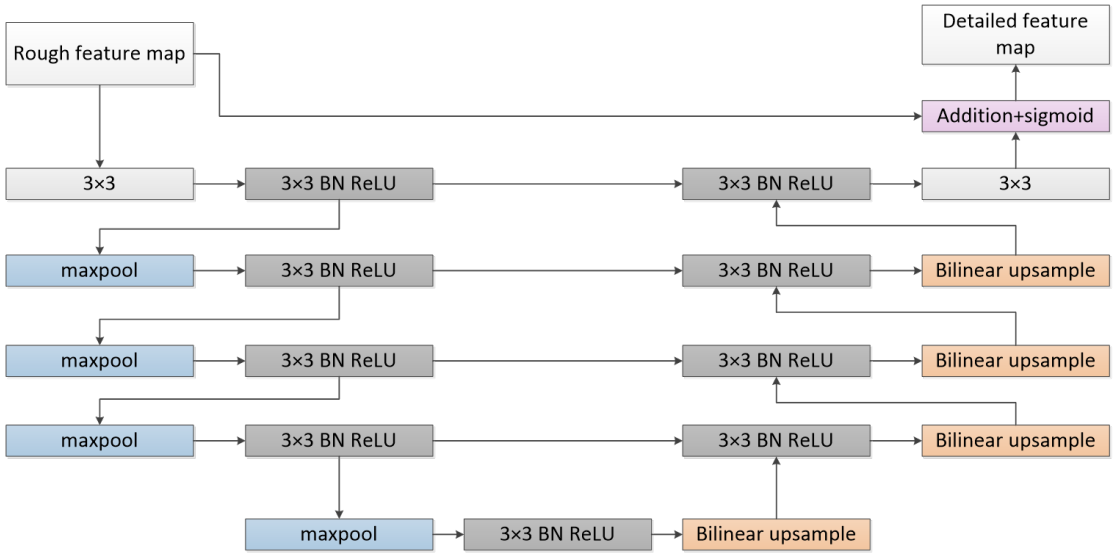


Figure 3: Residual refinement module of Qin et al.

2.4. Hybrid Loss Function

To ensure a comprehensive optimization of the objective during the training process and hence improve the segmentation results, we train the network using the following hybrid function:

$$L_S = \alpha L_{bce} + \beta L_{dice} + \gamma L_{iou} \tag{1}$$

where L_{bce} is the binary cross-entropy (*BCE*) loss, L_{dice} is the Dice coefficient loss, L_{iou} is the intersection-over-union (*IoU*) loss, and α, β, and γ are weighting coefficients that balance the contribution of each loss.

The *BCE*, Dice, and *IoU* losses track the accuracy of classification at the pixel level, the accuracy of segmentation of the target region, and the spatial consistency of the overall segmented region, respectively. Combining these three losses balances the classification accuracy, segmentation accuracy, and spatial consistency during training.

2.5. Architecture of the MBR–UNet Network

The model in this paper adopts UNet-like U-shaped codecs and a hopping connection structure. The model framework is shown in Figure 4. The left (coding) path of MBR–UNet contains five layers of Res2Net block modules and four downsampling layers for feature extraction. Downsampling captures the high-level semantic features of the input image, improving the network's understanding and processing of the image. By installing a Res2Net block module as a convolutional layer, we mitigate the problematic gradient vanishing during the downsampling process; moreover, the unique sub-path structure of the Res2Net block improves the capability of the network feature representation. The right (decoding) path consists of two parts. The first part containing four layers of MSA–SPP modules and four upsampling layers reduces the high-level semantic features extracted in the coding part to the resolution of the original image through upsampling and feature fusion, finally generating pixel-level segmentation predictions. The second part upsamples the feature maps at each stage to match the size of the original map and accumulates them into a segmentation probability map, which is further optimized for edge segmentation through the final boundary refinement module, thus improving the generalization of the model.

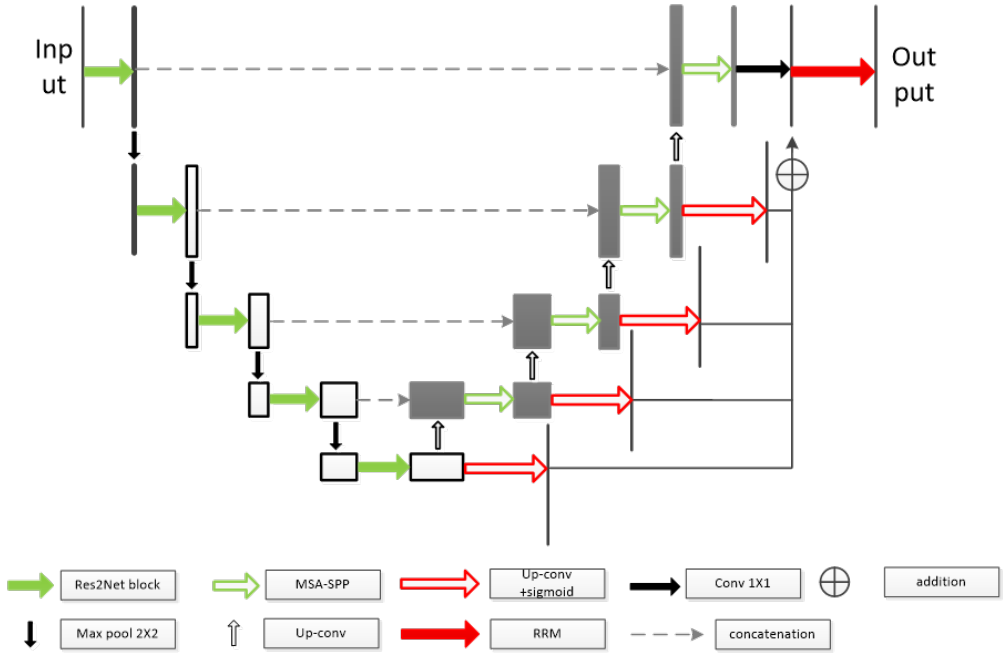


Figure 4: Structure of the Multiscale fusion and Boundary Refinement UNet (MBR–UNet) model.

3. Experimental Setup and Training

3.1. Dataset Settings

Experiments were performed on the lung segmentation dataset provided in the 2017 Kaggle competition. The dataset contains 267 two-dimensional lung image slices and corresponding lung parenchyma mask images. The resolution of each lung solid image and mask image is 512 $\times$

512. To increase the diversity of data samples, additional samples were generated through data enhancement approaches (horizontal flips and rotations through different angles).

3.2. Experimental Environment Setting

The hardware environment of all experiments was configured as shown in Table 1.

Table 1: Experimental environment.

Experimental platforms	Environmental Configuration
Operating system	Win10
Deep learning frameworks	Pytorch-gpu1.13.1
Programming language	Python3.7
CPU	i5-12600KF
GPU	RTX 3070 Ti
CUDA	NVDIA CUDA 11.7

The hyperparameters were set up as follows: batch size = 8; number of epochs = 200. The network parameters were updated with an Adams optimizer. The model was trained using the early stopping method to limit the number of training batches and hence prevent overfitting. The training was ceased if the model accuracy remained unchanged through 20 consecutive training batches.

3.3. Evaluation Indicators

The model was evaluated in terms of four segmentation-evaluation metrics: Accuracy (*ACC*), Dice similarity coefficient (*DSC*), IoU, and Hausdorff distance (*HD*). *ACC* reflects the ability of the model to predict the pixel-level accuracy. The *DSC* is a widely used semantic-segmentation indicator that measure the matching degree between the predicted and true segmentations. The *IoU*, which measures the overlap region between the predicted and real segmentation, is a key metric of segmentation effectiveness. The *HD* quantifies the maximum distance between the predicted and true boundaries, with smaller values indicating tighter boundary alignment. The *ACC*, *DSC*, *IOU*, and *HD* are respectively calculated as follows and the variables are defined in Table 2:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$\mathrm{DSC} = \frac{2TP}{2TP + FP + FN} \tag{3}$$

$$IoU = \frac{|\, A \cap B \,|}{|\, A \cup B \,|} \tag{4}$$

$$HD = max\left\{max_{a \in A} min_{b \in B} d(a,b), max_{a \in A} min_{b \in B} d(b,a)\right\} \tag{5}$$

4. Analysis of Experimental Results

4.1. Comparative Experiments of Different Segmentation Models

To demonstrate whether the proposed segmentation model can effectively enhance the segmentation of lung parenchyma, the performance indices of the MBR–UNet model were compared with those of four classical models, namely, UNet, DeepLabv3, UNet++ (Zhou et al., 2019), and U2Net (Qin et al., 2020), on the same dataset in the same environmental configuration. The comparison results are presented in Table 3.

Table 2: Definitions of variables in the evaluation indices (Eqs. (2)–(5)).

Variable	Interpretation
TP	Number of pixels correctly predicted in the positive category
TN	Number of pixels correctly predicted in the negative category
FP	Number of pixels in the negative category which are incorrectly assigned to the positive category
FN	Number of pixels in the positive category that are incorrectly assigned to the negative category
A	Predicted set of pixels in the segmentation result
B	A collection of pixels in the real segmentation result
d(a, b)	Distance from points in set *a* to points in set *b*

Table 3: Comparison of results of the evaluated segmentation models.

Models	*DSC* (%)	*IoU* (%)	*ACC* (%)	*HD* (mm)
UNet	93.85	93.11	96.84	25.92
DeepLabv3	95.56	94.75	98.23	20.05
UNet++	94.67	93.66	97.78	22.38
U2Net	95.37	95.12	98.15	18.96
MBR–UNet	97.35	96.34	98.79	12.82

As confirmed in Table 3, the segmentation performance of the proposed network model surpasses those of the other methods by different degrees. MBR–UNet improved the DSC, IoU, and ACC by 3.50%, 3.23% and 1.95%, respectively, from those of UNet network. Moreover, the HD of MBR–UNet was notably reduced from those of the other models and was nearly half that of the UNet network. The high HD performance was attributed to the RRM module, which optimizes the boundary defects and largely improves the boundary identification.

To show the final segmentation effect of each segmentation model, three pictures were randomly selected from the test set and the segmentation results of the models were visualized as shown in Figure 5. As evidenced in the figure, the classical models did not completely remove the influences of irrelevant regions such as blood vessels and bones. They were also prone to over-segmentation and under-segmentation and produced relatively rough segmentations of the lung parenchyma boundaries. The segmented lung parenchyma region of the proposed method best resembled the real mask, demonstrating that the MBR–UNet segmentation model effectively reduced the over-segmentation and under-segmentation problems and produced a refined segmentation of the lung boundary.

4.2. Ablation Experiment

The contribution of each component of the segmentation network to the segmentation performance of each model was evaluated through ablation experiments. The results are tabulated in Table 4.

Introducing the Res2Net block and the MSA–SPP module noticeably improved the *DSC*, *IoU*, and *ACC* metrics because the Res2Net block effectively extracts features through its multiscale paths, whereas the MSA–SPP module captures the fine-grained features and effectively integrates the multiscale contextual information, thus optimizing the model's retention of the feature information extracted from the coded part. Meanwhile, the RRM module considerably reduced the *HD* metric, confirming the efficacy of the boundary refinement module. The segmentation performance was further improved by adopting the hybrid loss, which calculates the losses at different depth levels.

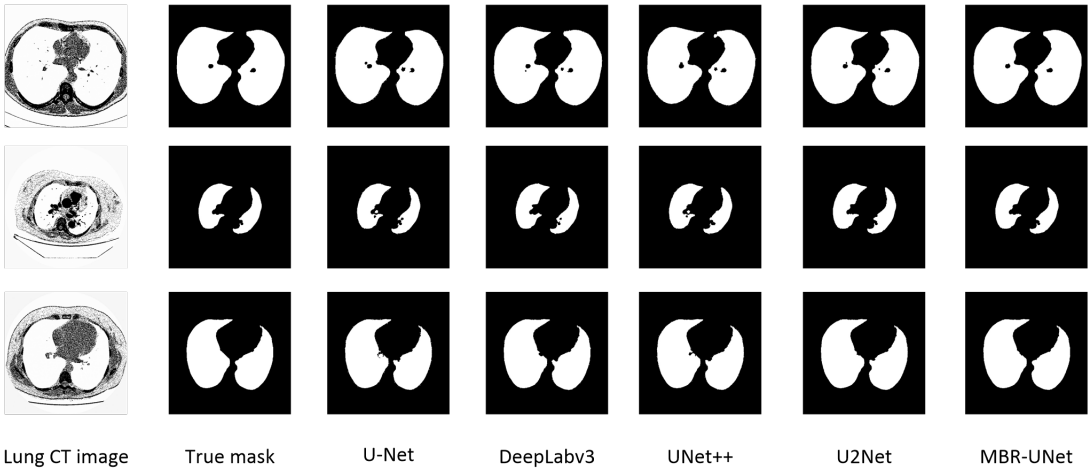


Figure 5: Segmentation effects of the evaluated segmentation models.

Table 4: Results of the ablation experiments.

Models	*DSC* (%)	*IoU* (%)	*ACC* (%)	*HD* (mm)
UNet	93.85	93.11	96.84	25.92
UNet & Res2Net block	95.21	94.45	97.67	22.42
UNet & Res2Net block & MSA–SPP	96.38	95.36	98.13	19.95
UNet & Res2Net block & MSA–SPP & RRM	96.59	95.73	98.34	14.63
UNet & Res2Net block & MSA–SPP & RRM & Hybrid loss	97.35	96.34	98.79	12.82

5. Summary

We have designed a lung parenchyma segmentation network model based on a U-shaped coding and decoding structure. The Res2Net block module performs feature extraction along the coding path and an MSA–SPP module is designed for feature fusion along the decoding path and boundary refinement in the output results. Finally, the model training is optimized using the hybrid loss function, which is expected to improve the accuracy of lung parenchyma segmentation by ensuring

adequate feature extraction and fusion and refining the segmentation boundary. Experiments on the lung segmentation dataset of the Kaggle competition demonstrated the advantages of the proposed model in lung parenchyma segmentation; in particular, the model more completely segmented the lung parenchyma region than the classical methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62201231) and the China Postdoctoral Science Foundation (Grant No. 2022MD713827).

References

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

Zijun Deng, Xiaowei Hu, Lei Zhu, Xuemiao Xu, Jing Qin, Guoqiang Han, and Pheng-Ann Heng. R3net: Recurrent residual refinement network for saliency detection. In *Proceedings of the 27th international joint conference on artificial intelligence, Menlo Park, CA, USA*, volume 684690. AAAI Press Menlo Park, CA, USA, 2018.

Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662, 2019.

Lei Geng, Siqi Zhang, Jun Tong, and Zhitao Xiao. Lung segmentation method with dilated convolution based on vgg-16 network. *Computer Assisted Surgery*, 24(sup2):27–33, 2019.

Meng-Hao Guo, Cheng-Ze Lu, Qibin Hou, Zhengning Liu, Ming-Ming Cheng, and Shi-Min Hu. Segnext: Rethinking convolutional attention design for semantic segmentation. *Advances in Neural Information Processing Systems*, 35:1140–1156, 2022.

Anita Khanna, Narendra D Londhe, S Gupta, and Ashish Semwal. A deep residual u-net convolutional neural network for automated lung segmentation in computed tomography images. *Biocybernetics and Biomedical Engineering*, 40(3):1314–1327, 2020.

Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern recognition*, 106:107404, 2020.

Xuebin Qin, Deng-Ping Fan, Chenyang Huang, Cyril Diagne, Zichen Zhang, Adrià Cabeza Sant'Anna, Albert Suarez, Martin Jagersand, and Ling Shao. Boundary-aware segmentation network for mobile and web applications. *arXiv preprint arXiv:2101.04704*, 2021.

Brahim Ait Skourt, Abdelhamid El Hassani, and Aicha Majda. Lung ct image segmentation using deep neural networks. *Procedia Computer Science*, 127:109–113, 2018.

Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.

Lun Zhu, Yinghui Cai, Jiahao Liao, and Fan Wu. Lung parenchyma segmentation based on u-net fused with shape stream. *IEEE Access*, 2024.