EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Fast Lung Image Segmentation Using Lightweight VAEL-Unet

Xiulan Hao[1,*], Chuanjin Zhang[1,†],Shiluo Xu[1,‡]

[1]Institute for Artificial Intelligence, School of Information Engineering, Huzhou University, Zhejiang, China

Abstract

INTRODUCTION: A lightweght lung image segmentation model was explored. It was with fast speed and low resouces consumed while the accuracy was comparable to those SOAT models.
OBJECTIVES: To improve the segmentation accuracy and computational efficiency of the model in extracting lung regions from chest X-ray images, a lightweight segmentation model enhanced with a visual attention mechanism called VAEL-Unet, was proposed.
METHODS: Firstly, the bneck module from the MobileNetV3 network was employed to replace the convolutional and pooling operations at different positions in the U-Net encoder, enabling the model to extract deeper-level features while reducing complexity and parameters. Secondly, an attention module was introduced during feature fusion, where the processed feature maps were sequentially fused with the corresponding positions in the decoder to obtain the segmented image.
RESULTS: On ChestXray, the accuracy of VAEL-Unet improves from 97.37% in the traditional U-Net network to 97.69%, while the F1-score increases by 0.67%, 0.77%, 0.61%, and 1.03% compared to U-Net, SegNet, Res-Unet and DeepLabV3+ networks. respectively. On LUNA dataset. the F1-score demonstrates improvements of 0.51%, 0.48%, 0.22% and 0.46%, respectively, while the accuracy has increased from 97.78% in the traditional U-Net model to 98.08% in the VAEL-Unet model. The training time of the VAEL-Unet is much less compared to other models. The number of parameters of VAEL-Unet is only 1.1M, significantly less than 32M of U-Net, 29M of SegNet, 48M of Res-Unet, 5.8M of DeeplabV3+ and 41M of DeepLabV3Plus_ResNet50.
CONCLUSION: These results indicate that VAEL-Unet's segmentation performance is slightly better than other referenced models while its training time and parameters are much less.

Received on 7 January 2024; accepted on 05 April 2024; published on 08 April 2024

Keywords: Medical image segmentation, Deep learning, VAEL-Unet, Attention module

Copyright © 2024 X. Hao *et al.*, licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi:10.4108/eetsis.4788

1. Introduction

In [1], trends in scientific production on artificial intelligence and health in Latin America in Scopus was explored. The conclusion was that main topics were predictive models and the application of artificial intelligence for classifying, diagnosing and treating diseases .

When diagnosing respiratory diseases, doctors must meticulously differentiate the lung regions in chest X-ray images and exclude the influence of other areas. This can result in reduced efficiency and fatigue in doctors, increasing the risk of misdiagnosis.

To attack these problems, medical image processing algorithms based on deep learning have emerged as a prominent solution [2], and they are also used in other assisted diagnosis, such as ophthalmic diseases diagnosis in [3–6] and diabetic eye disease identification in [7, 8], etc. Deep learning is also used in other scientific research, such as natural language processing [9], remote sensing image processing [10], etc.

Some researchers devoted to identifying neurological disorders from EEG signals, early detection of mild cognitive impairment [11–14]. Some also attemped to identify antisocial behavior [15], and automatically

*Email: xiulanhao@fudan.edu.cn
†Chuanjin Zhang contributed equally to this work as the co-first author.
‡Corresponding author. Email: xushiluo@zjhu.edu.cn

detect autism spectrum disorder from EEG [16]. Mental health was analyzed based on emotion recognition from facial expressions and psychometric evaluations [17]. A personalized arrhythmia detection system based on attention mechanism called personAD, was proposed in [18]. On MIT-BIH Arrhythmia Database, the arrhythmia detection system achieved 98.03%.

Breast lesions were automatically and accurately segmented using max flow and min cut problems in the continuous domain over phase preserved denoised images in [19].

Fully convolutional network (FCN) network was introduced in [20], which utilizes convolutional layers for feature extraction and downsampling in the spatial dimension, followed by transposed convolutional layers for upsampling and feature fusion. This results in predictions of the same size as the input image, effectively addressing semantic-level image segmentation. However, the FCN network lacks cross-layer feature fusion, leading to lower segmentation accuracy.

Consequently, U-Net network was proposed in [21], which builds upon the FCN network by incorporating a cross-layer feature fusion mechanism in the encoder [22]. This mechanism involves concatenating high-level features with low-level features to enhance feature expression. The U-shaped structure of U-Net facilitates the capture of contextual information at various scales and thereby improving segmentation accuracy.

To further enhance image segmentation accuracy, Res-Unet network was adopted in [23], which combines residual learning with the U-Net network. The Res-Unet network demonstrates superior accuracy and robustness in image segmentation, particularly in contour segmentation. An improved U-Net neural network for the auxiliary diagnosis of intracerebral hemorrhage was proposed [24], which realizes the automatic segmentation of the hemorrhage on CT images. In [25], the performance of Multi-scale Fusion Attention U-Net (MSFA-U-Net) in thyroid gland segmentation on localized computed tomography (CT) images for radiotherapy was explored . In [26], MBConvBlock encoder module was adopted, decoder module was interpolated and reconstructed and triple threshold strategy was used to improve U-Net network and good performance was achieved in pneumothorax X-ray image segmentation.

This year, an enhanced medical image segmentation model, RAAU-Net, based on the U-Net architecture was proposed in [27]. It has better performance across all metrics in the segmentation tasks of 2D medical images, including retinal nerves, skin lesions, and lung regions. A novel lung parenchyma segmentation network named ACX-UNet was proposed in [28], which incorporates attention mechanisms and cyclic cross-feature extraction strategies. It produces prediction maps that more closely resemble the true labels. In [29], a new segmentation mechanism was introduced, which is based on Fuzzy C-Means (FCM) and various features by incorporating dual FCM. Compared to several previous methods, this approach resulted in improvements of 4.2210% and 2.3150% in the Jaccard Index and Dice Coefficient, respectively.

In lung region image segmentation based on chest X-ray images, main challenges include the complexity of anatomical structures within images, such as the lungs and ribs. These structures exhibit close grayscale values in X-ray images, especially at the junction of the lungs and ribs. Their visual similarity makes it difficult for automatic segmentation algorithms to distinguish. In addition, the quality of chest X-ray images is affected by many factors, such as the angle of imaging, patient body size, and radiation dose, which leads to huge differences in contrast, brightness, and clarity in the images, making accurate segmentation challenging. At the same time, high-precision segmentation models often have many parameters, requiring long training time and large computing resources, which limits their applications in resource-constrained situations. In some resource-constrained medical settings, there may be a lack of high-performance computing resources. How to maintain segmentation accuracy while reducing model resource consumption becomes an important challenge.

Therefore, to compromise between high accuracy and low resource consumption, U-Net is used as the benchmark and integrated with the lightweight MobileNetV3(Small) network, along with the CBAM (Convolutional Block Attention Module). This integration leads to the development of a vision and attention enhanced lightweight lung region image segmentation model, named VAEL-Unet.

The main contributions of this work are:

1) VAEL-Unet is designed, which is a lightweight lung region segmentation model incorporating visual attention enhancement;

2) The CBAM attention mechanism is introduced in VAEL-Unet to adaptively learn the importance of different regions in the image and concentrate attention on the more informative areas;

3) Experimental results on benchmark datasets show that VAEL-Unet can slightly improve the segmentation performance while training time and parameters are reduced sharply.

2. Methods

The VAEL-Unet consists of three components: an encoder, a decoder, and a feature fusion module. The encoder employs the lightweight MobileNetV3 for feature extraction, the decoder utilizes that of the U-Net. In the feature fusion module, the CBAM attention mechanism is introduced to extract useful

features, followed by feature concatenation to enhance the segmentation accuracy of the model.

2.1. Related Models

U-Net network model. U-Net demonstrates exceptional performance in medical image segmentation tasks [30] . The model comprises symmetric encoder and decoder components. The encoder consists of multiple convolutional layers and pooling layers that extract high-level feature representations from the input image.

The decoder incorporates multiple upsampling layers and convolutional layers, to expand the feature maps generated by the encoder to match the dimensions of the original input image and generating pixel-level segmentation masks [31]. To enhance the quality of segmentation, U-Net incorporates skip connections to connect feature maps from corresponding layers of the encoder and decoder. This makes full use of both low-level and high-level features.

MobileNetV3 network model. MobileNetV3 network model is designed to address computational limitations in real-world scenarios. It is commonly used as feature extraction component in various models to reduce parameters and improve training speed [32].

The feature extraction component of the MobileNetV3 is predominantly composed of bottleneck (bneck) modules. The bneck module, by integrating technologies such as depthwise separable convolution [33] , inverted residual structures [34], and the Squeeze-and-Excitation mechanism [35]. and the efficient H-swish activation function, achieves the goal of enhancing model performance while reducing computational complexity and the number of parameters. The architecture is illustrated in Figure 1.

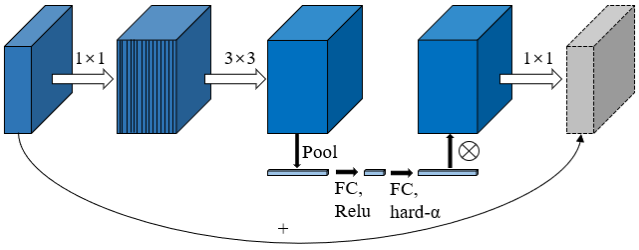


Figure 1. Bneck Module Architecture

Depthwise separable convolution consists of two steps: depthwise convolution and pointwise convolution. The parameters are calculated as follows:

$$P_{DC} = m \times k \times k \tag{1}$$

$$P_{PC} = m \times n \times 1 \times 1 \tag{2}$$

$$P = P_{DC} + P_{PC} \tag{3}$$

$$R = \frac{m \times k \times k + m \times n \times 1 \times 1}{m \times n \times k \times k} = \frac{1}{n} + \frac{1}{k^2} \tag{4}$$

where $P_{DC}, P_{PC}$and P are the number of parameters for deep convolution, point-wise convolution, and depth-wise separable convolution, respectively, k is the size of the convolutional kernel, m is the number of input channels, and n is the number of output channels. R is the ratio between parameters of depth-wise separable convolution and that of conventional convolution. This ratio elucidates that the number of parameters for depth wise separable convolution is significantly lower than that for conventional convolution, thereby substantially reducing the training time of the network.

The inverted residual structure reverses the traditional residual structure process by first extending the input to a higher-dimensional feature space for processing, and then compressing it back to a lower dimension by a 1×1 convolutional layer. This structure effectively enhances the model's ability to process information while keeping a lower number of parameters. The Squeeze-and-Excitation mechanism dynamically adjusts the weights of channels by explicitly modeling the dependencies between channels, allowing the network to adaptively reinforce important features and suppress unimportant ones, thereby improving model accuracy and generalization ability. H-swish, a variant of the Swish activation function, introduces a hard gating to simplify the computation of Swish. While keeping nonlinear characteristics, it enhances the network's learning capability and efficiency.

The integration of these technologies allows the bneck module to offer outstanding performance within a lightweight network architecture, optimizing the utilization of computational resources while keeping or improving the accuracy of the model. This is important for deep learning models running in computationally constrained settings.

CBAM attention module. The CBAM is an efficient and lightweight attention module that can be integrated into any convolutional neural network architecture and trained end-to-end with the base network [36]. It focuses on both channel and spatial attention and achieves better performance compared to attention modules that only focus on a single aspect [37].

The CBAM module is divided into Channel Attention Module and Spatial Attention Module [38]. The mathematical expression for CBAM is shown in Equation (5),

$$\begin{aligned} F' &= M_c(F) \otimes F, \\ F'' &= M_s(F') \otimes F' \end{aligned} \tag{5}$$

where $\otimes$ denotes element-wise multiplication, F is the input feature map, $M_c(F)$ is the channel attention map output from the channel attention module,$M_s(F')$ is the spatial attention map output from the spatial attention

module, and F'' is the final output feature map of the CBAM attention module.

Channel attention module. It manipulates the channel information of input images to emphasize essential channel features and suppress insignificant ones.

Input feature F is pooled globally based on its width and length, and global average pooling and maximum pooling features are formed. The pooled features are fed into a multi-layer perceptron (MLP) [39].

The computation of channel attention map is described by Equation (6),

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) \\ &\quad + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max}))) \end{aligned} \tag{6}$$

where σ is the Sigmoid activation function, W_0 and W_1 are the weights of the MLP, $W_0 \in R^{C/r \times C}$, $W_1 \in R^{C \times C/r}$, and r is the dimension reduction factor.

Spatial attention module. It is employed to modify the spatial information of the input image, emphasizing crucial spatial locations while suppressing irrelevant ones.

Feature map F' is its input; F^s_{avg} and F^s_{max} are channel-wise global max pooling and global average pooling.

The computation of spatial attention map $M_s(F')$ is defined by Equation (7),

$$\begin{aligned} M_s(F') &= \sigma(f^{7\times7}([AvgPool\,(F');MaxPool(F')])) \\ &= \sigma(f^{7\times7}([F^s_{avg};F^s_{max})) \end{aligned} \tag{7}$$

where σ is the Sigmoid activation function, $f^{7\times7}$ is the convolution operation with a 7×7 kernel size.

2.2. VAEL-Unet architecture

Figure 2 illustrates the overall architecture of the VAEL-Unet segmentation model. The entire process of the VAEL-Unet model is described by equations (8) and (9),

$$\begin{aligned} x_i &= f_{enc}\,(x) \\ &= b_i\,(b_{i-1}\,(b_{i-2}\,(\cdots b_1\,(W\,(x)\cdots)))), \\ &\quad i = 1,2,3,4,5 \end{aligned} \tag{8}$$

$$\begin{aligned} y_j &= f_{dec}\left(y_{j+1}, x_j\right) \\ &= f_{act}\left(W\left(G_{up}\left(y_{j+1}\right) \oplus A\left(x_j\right)\right)\right), \\ &\quad j = 4,3,2,1 \end{aligned} \tag{9}$$

where x is the input image of the model, f_{enc} is the operation of the encoder, x_i is the output feature map of the layer i in the encoder, b_i is the operation of the bneck module in the layer i , W is the convolution operation; f_{dec} is the operation of the decoder, y_j is the output feature map of the layer j in the decoder, f_{act} is the activation function, usually the ReLU function, G_{up} is the upsampling operation, typically achieved through deconvolution or bilinear interpolation, $\oplus$ is the concatenation operation, A is the operation of the CBAM attention module.

3. Results

3.1. Dataset

The dataset used is the public Chest X-ray dataset from Kaggle. It consists of lung mask images and chest X-ray images. The mask images were in one-to-one correspondence with the chest X-ray images. According to the 8:1:1 ratio, 8000 samples in the Chest X-ray dataset were randomly selected as the training set, 1000 samples as the validation set, and 1000 samples as the test set to form the ChestXray dataset. To meet the image size requirements of segmentation network model, it is necessary to scale the image in equal proportion to achieve the required size.

To validate the generalization capability of the model, we utilized the widely-used LUNA16 dataset [40, 41]. Following an 8:1:1 ratio, 640 samples from the LUNA16 dataset were randomly selected as the training set, while 80 samples as the validation set and another 80 samples as the test set, forming the LUNA dataset. Since each sample in the LUNA dataset is a three-dimensional image, a slicing process was performed on each sample to extract slices from the exact middle of the sample image and the corresponding mask image. Furthermore, the dimensions of the slices were uniformly scaled to meet the requirements of the segmentation model's input.

3.2. Data Preprocessing

Due to different size of original images in the dataset, a uniform scaling process is applied in preprocessing, setting the dimensions of images to 256×256. Simultaneously, to further assure the robustness of the training model, data augmentation techniques such as random rotation, horizontal flipping, and vertical flipping are employed to expand the dataset. The processed data is illustrated in Figure 3.

3.3. Experiment setups

The experimental setups are shown in Table 1.The input image size for the model is set to 256×256. The Adam optimizer is employed to update the network's model weights, with a learning rate of 0.0001. Given the performance of GPU, image samples randomly drawn for each batch is set to 4. Epoch in training is fixed at 100, and the optimal network model obtained during training is reserved.

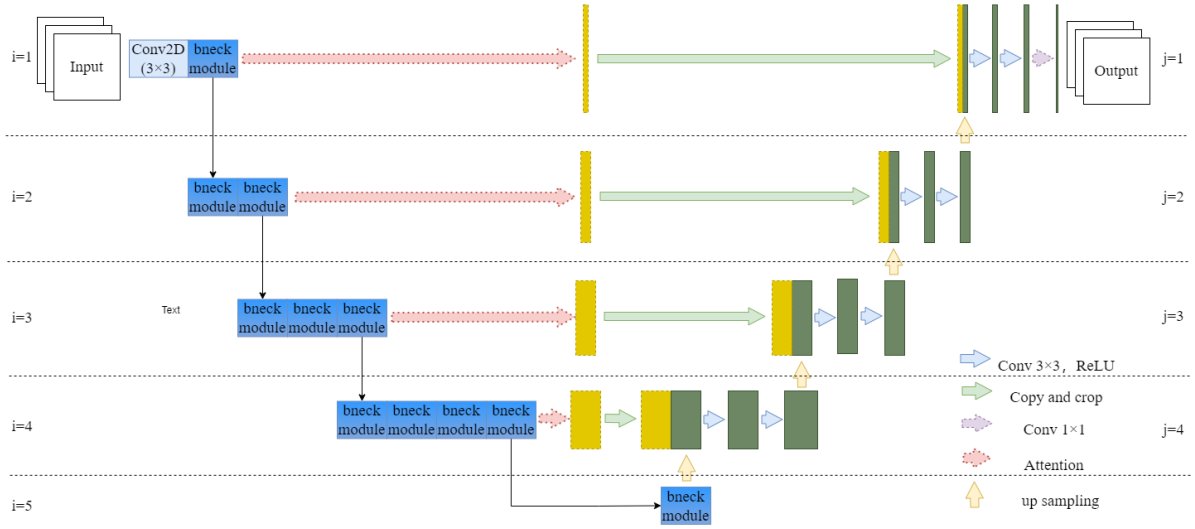


Figure 2. VAEL-Unet Architecture

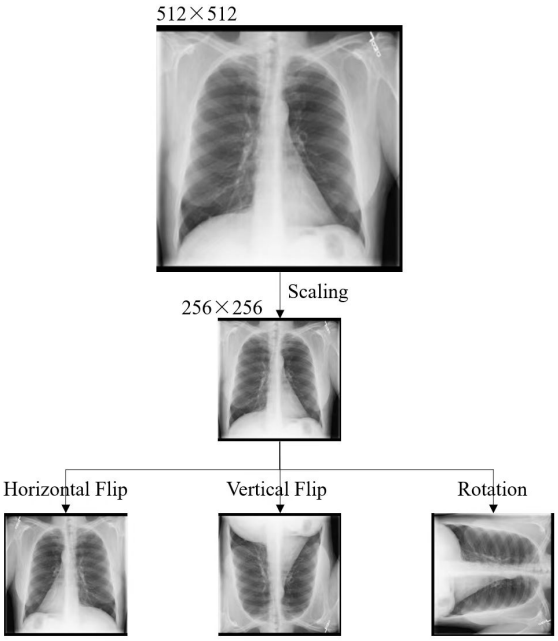


Figure 3. Image Preprocessing

The task of segmenting lung regions is a binary classification to determine whether a given pixel belonged to the lung region or not. The activation function utilized was the commonly used Sigmoid function [42]. The Dice loss [43] was used as the loss function, as shown in Formula (10),

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{n} p_i t_i + \varepsilon}{\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} t_i^2 + \varepsilon} \tag{10}$$

where p_i is the probability of the model's prediction for pixel i, t_i is the true label of pixel i, n is the total number of pixels, and ε is a very small value to avoid division by zero.

Table 1. Experiment setups

Parameters	Values
CPU	Intel Xeon Silver 4214 @ 2.20GHz, 12-core
Memory	DDR 64GB
GPU	NVIDIA Tesla V-100
Operating System	Linux
Programming language	Python 3.9.16
IDE	Pycharm Community Edition(2021.3)
Framework	PyTorch(2.0.0)

3.4. Evaluation metrics

The performance of the segmentation model is assessed using the following metrics: accuracy (A_c), recall , intersection over union (IoU), and F1-score (F_1). The specific calculations are defined by the following formulas,

$$A_c = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \times 100\% \tag{11}$$

$$Recall = \frac{T_P}{T_P + F_N} \times 100\% \tag{12}$$

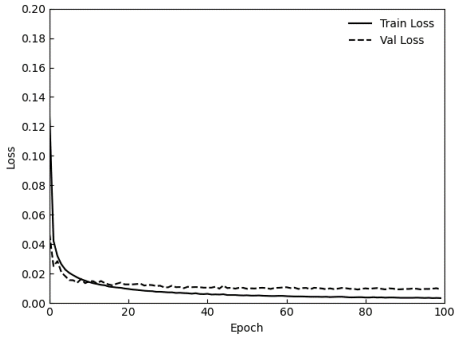
$$IoU = \frac{T_P}{T_P + F_P + F_N} \times 100\% \tag{13}$$

$$F_1 = 2\frac{T_P}{F_P + 2T_P + F_N} \times 100\% \tag{14}$$

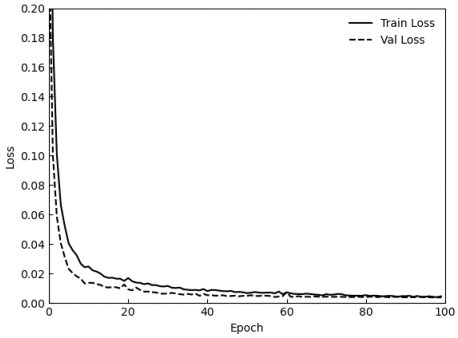
where T_P is true positives, F_N is false negatives, F_P is false positives, T_N is true negatives.

3.5. Model validation

The loss curves of the training are shown in Figure 4, where 4(a) is the loss curve on ChestXray, and 4(b) is the loss curve on LUNA. From both plots, it can be observed that the training and validation losses of the model decrease as the number of Epochs increases, eventually reaching a stable state. Furthermore, the losses in both cases are reduced to below 0.02.



(a) Training Loss Curve for ChestXray



(b) Training Loss Curve for LUNA

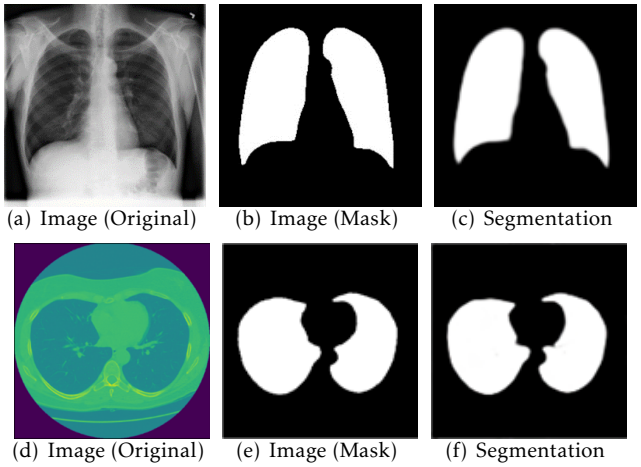
Figure 4. Training Loss Curve

3.6. Visualization of Segmentation

To more intuitively reflect the effect of the VAEL-Unet on lung region segmentation, two chest X-ray images were randomly selected from the test set for segmentation and display. The results are shown in Figure 5, where the white area is the segmented lung region, and the black part represents the area outside the lung. It can be seen from the figure that VAEL-Unet can segment the lung area from the chest X-ray image very well.

3.7. Comparisons among different models

VAEL-Unet is compared with the U-Net, SegNet [44], Res-Unet [45], DeepLabV3+ [46] and DeepLabV3Plus_ResNet50 [47] on ChestXray and LUNA. The results are compared in A_c, Recall, IoU, F_1, training time and parameters. In the following figures, DeepLabV3Plus_ResNet50 is abbreviated as DL+ResNet50.



(a) Image (Original) (b) Image (Mask) (c) Segmentation

(d) Image (Original) (e) Image (Mask) (f) Segmentation

Figure 5. Visualization of Segmentation for Chest X-ray (a-c) and LUNA Images (d-f)

By analyzing Figures 6(a) and 7(a), it is evident that the training time of the VAEL-Unet segmentation model is less compared to the U-Net, SegNet, Res-UnetDeepLabV3+, and DeepLabV3Plus_ResNet50.

From Figure 6(b), it can be observed that VAEL-Unet outperforms U-Net, SegNet, Res-Unet, and DeepLabV3+ models on ChestXray in A_c, IoU, F_1, achieving 97.69%, 93.65%, and 94.85%, respectively. While the Res-Unet network model exhibits a deeper network architecture and stronger feature extraction capability, its sensitivity to interfering pixel points leads to moderate precision but higher recall . Among the compared models, Res-Unet attains the highest recall, with F1-score of 94.24%.

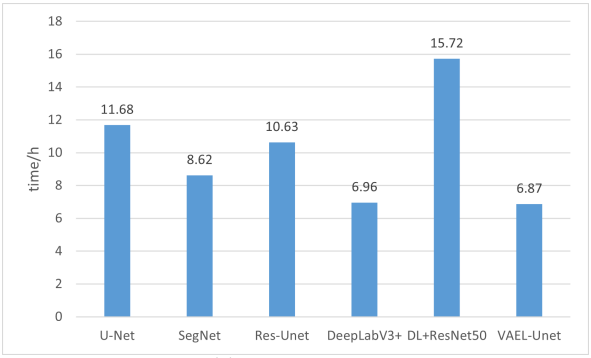
The accuracy of VAEL-Unet improved from the conventional U-Net network's 97.37% to 97.69%. When compared with U-Net, SegNet, Res-Unet, and DeepLabV3+ networks, the F1-score experiences an increment of 0.67%, 0.77%, 0.61%, and 1.03%, respectively.

As shown in Figure 7(b), that the VAEL-Unet outperforms the traditional U-Net, SegNet, Res-Unet and DeepLabV3+ on LUNA. The F1-score demonstrates improvements of 0.51%, 0.48%, 0.22% and 0.46%, respectively, while the accuracy has increased from 97.78% in the traditional U-Net to 98.08% in the VAEL-Unet These results indicate that the VAEL-Unet exhibits strong generalization capability and consistently delivers excellent performance across different datasets.

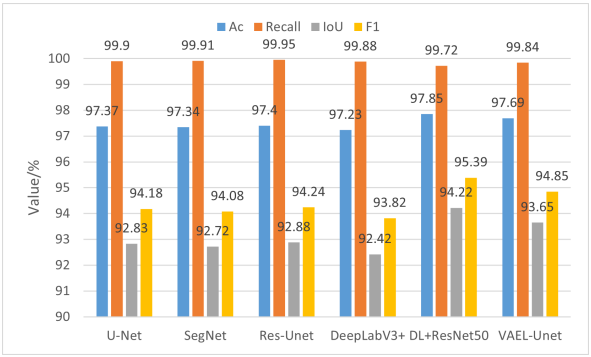
4. Discussion

4.1. Ablation Analysis

To assess the effectiveness of combining MobileNetV3 and CBAM within the VAEL-Unet framework, a series of ablation experiments were done. These

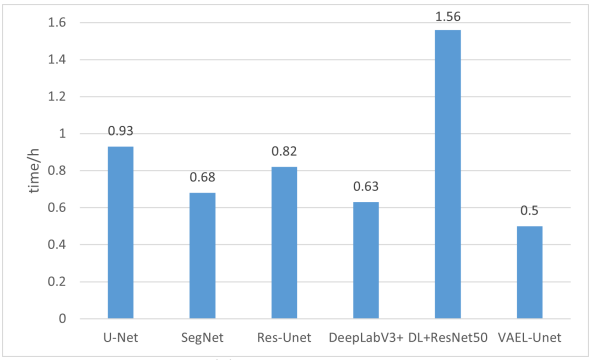


(a) Training time

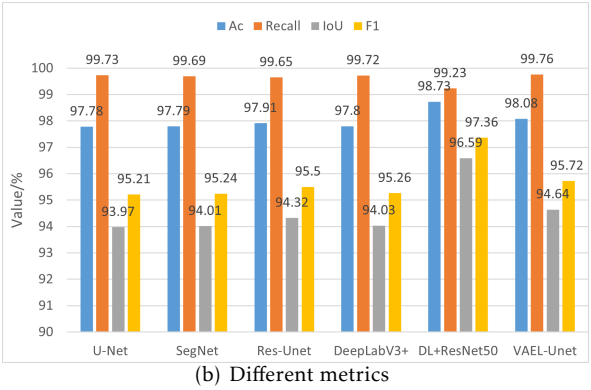


(b) Different metrics

Figure 6. Comparison for Chest X-ray images



(a) Training time



(b) Different metrics

Figure 7. Comparison for LUNA images

experiments evaluated the performance of the models under different ablation settings, focusing on metrics including A_c, Recall, IoU, F_1. The results of the model ablation experiments are presented in Table 2. In this table, "–" signifies the module is excluded, while "✓" signifies the module is included.

It can be observed that the proposed VAEL-Unet network model exhibits improvements in accuracy, IoU, and F1 compared to the U-Net model, U-Net model with the CBAM attention module, and the U-Net model combined with MobileNetV3 without the CBAM attention module, across both the ChestXray and LUNA datasets. Therefore, it can be inferred that the VAEL-Unet effectively enhances the performance of pulmonary region segmentation from chest X-ray images. This justifies incorporating MobileNetV3 and CBAM within the VAEL-Unet framework to enhance segmentation performance.

4.2. Limitations of the conventional U-Net network model

The U-Net network model requires effective extraction of abstract features from low-resolution, blurry boundary, and poor contrast chest X-ray images to enhance the segmentation performance. Although the U-Net network model employs an encoder-decoder architecture and skip connections to efficiently utilize features at different levels, it still has several limitations:

1) The extensive use of convolutional layers and pooling layers in the U-Net network model can effectively extract features at various scales. However, the use of conventional convolution operations may lead to information loss and overfitting, making it challenging to apply them to other image segmentation tasks [48].

2) The U-Net network model has a relatively shallow network depth [49], and typically layers range from 7 to 10 . This limitation may restrict the model's expressive capacity and hinder the accurate segmentation of chest X-ray images. Additionally, the limited number of network layers reduces the opportunities for feature learning, potentially resulting in the inability to capture richer feature representations.

3) The U-Net network model has a large number of parameters, leading to longer training times. This limits its application in real-time scenarios.

4.3. Advantages of VAEL-Unet model

To compromise between model performance and computing resource consumption, we propose VAEL-Unet. The bneck modules of the lightweight MobileNetV3 (Small) were employed as part of the encoder of the VAEL-Unet to extract deep features. When these features are fused with each layer of the decoder, the CBAM attention module extracts expressive features initially, followed by feature splicing. This approach

Table 2. Ablation experiments on ChestXray and LUNA

Dataset	U-Net	MobileNetV3	CBAM	A_c/%	Recall/%	IoU/%	F_1/%
ChestXray	✓	–	–	97.37	99.90	92.83	94.18
	✓	–	✓	97.39	99.72	92.81	94.09
	✓	✓	–	97.47	99.88	93.11	94.42
	✓	✓	✓	97.69	99.84	93.65	94.85
LUNA	✓	–	–	97.78	99.73	93.97	95.21
	✓	–	✓	97.78	99.72	93.99	95.09
	✓	✓	–	97.94	99.85	94.29	95.43
	✓	✓	✓	98.08	99.76	94.64	95.72

eliminates the impact of invalid features on the model and resolves the challenging issue of improvement in segmentation accuracy to some degree. Therefore, the performance and parameters of the VAEL-Unet segmentation model are optimized.

VAEL-Unet effectively reduces the network's depth and parameters, thereby mitigating the issue of gradient explosion. It only has 1.1M parameters, significantly smaller than 32M of U-Net, 29M of SegNet, 48M of Res-Unet, 5.8M of DeepLabV3+ and and 41M of DeepLabV3Plus_ResNet50, as shown in Figure 8.

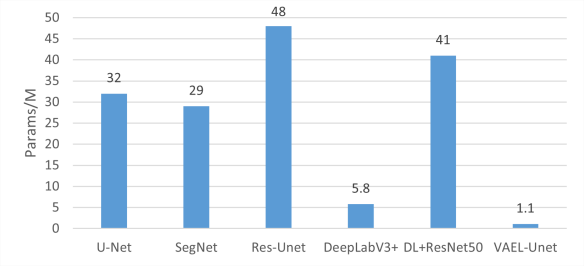


Figure 8. Comparison of Parameters

This reduction in parameters greatly alleviates the computational burden of the model. Consequently, the VAEL-Unet excels in model performance, parameters, and training time when compared to the traditional U-Net, SegNet, Res-Unet and DeepLabV3+.

Despite the slight decrease in accuracy, IoU, and F1 compared to the latest DeepLabV3Plus_ResNet50, VAEL-Unet needs less training time and parameters. The training time is reduced by approximately 56%, and the number of parameters is decreased by about 40 times compared to DeepLabV3Plus_ResNet50. From perspective of computational resources consumption, VAEL-Unet is better than DeepLabV3Plus_ResNet50. VAEL-Unet is a lightweight model, while DeepLabV3Plus_ResNet50 is a heavyweight model which gets better results at the cost of more time and space resources.

5. Conclusion

To improve the accuracy and efficiency of lung region extraction from chest X-ray images, we use U-Net as the baseline, combine lightweight MobileNetV3 (Small) and introduce the CBAM attention module to design a visually enhanced and lightweight lung region image segmentation model VAEL-Unet. VAEL-Unet utilizes MobileNetV3 for extracting abstract features, incorporates feature fusion from U-Net, and leverages the CBAM attention module to enhance informative features. This approach enhances the segmentation performance while reducing the model's parameters. Compared to other commonly used segmentation network models, VAEL-Unet achieves improved segmentation accuracy, significant enhancement in training time and reduction in parameters. Thus, it achieves better segmentation results by balancing recognition accuracy and computational efficiency.

CRediT authorship contribution statement

Xiulan Hao analyzed and interpreted the LUNA data, polished English and was a major contributor in writing and editing the manuscript. Chuanjin Zhang carried on the experiments, analyzed and interpreted the chestXray data, and was a major contributor in writing the manuscript. Shiluo Xu investigated the related work and polished English. All authors read and approved the final manuscript.

Declarations of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

[ChestXray] is available at:
https://www.kaggle.com/datasets/newra008/lung-mask-image-dataset

Funding

None.

Acknowledgments

Thanks editors and anonymous reviewers for their suggestions to improve this work.

References

[1] Gonzalez-Argote, J., Alonso-Galbán, P., Vitón-Castillo, A.A., Lepez, C.O., Castillo-Gonzalez, W., Bonardi, M.C. and Cano, C.A.G. (2023) Trends in scientific output on artificial intelligence and health in latin america in scopus. *EAI Endorsed Transactions on Scalable Information Systems* **10**(4): e5–e5. doi:http://dx.doi.org/10.4108/eetsis.vi.3231.

[2] Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H. and Nandi, A.K. (2022) Medical image segmentation using deep learning: A survey. *IET Image Processing* **16**(5): 1243–1267. doi:https://doi.org/10.1049/ipr2.12419, URL https://doi.org/10.1049/ipr2.12419.

[3] Wu, M., Lu, Y., Hong, X., Zhang, J., Zheng, B., Zhu, S., Chen, N. *et al.* (2022) Classification of dry and wet macular degeneration based on the convnext model. *Frontiers in Computational Neuroscience* **16**. doi:10.3389/fncom.2022.1079155, URL https://www.frontiersin.org/articles/10.3389/fncom.2022.1079155.

[4] Zhu, S., Zhan, H., Yan, Z., Wu, M., Zheng, B., Xu, S., Jiang, Q. *et al.* (2023) Prediction of spherical equivalent refraction and axial length in children based on machine learning. *Indian Journal of Ophthalmology* **71**(5): 2115–2131. doi:https://doi.org/10.4103/IJO.IJO_2989_22, URL https://doi.org/10.4103/IJO.IJO_2989_22.

[5] Zhu, S., Lu, B., Wang, C., Wu, M., Zheng, B., Jiang, Q., Wei, R. *et al.* (2022) Screening of common retinal diseases using six-category models based on efficientnet. *Frontiers in Medicine* **9**. doi:10.3389/fmed.2022.808402, URL https://www.frontiersin.org/articles/10.3389/fmed.2022.808402.

[6] Zheng, B., Liu, Y., He, K., Wu, M., Jin, L., Jiang, Q., Zhu, S. *et al.* (2021) Research on an intelligent lightweight-assisted pterygium diagnosis model based on anterior segment images. *Disease Markers* **2021**: 7651462. doi:https://doi.org/10.1155/2021/7651462, URL https://doi.org/10.1155/2021/7651462.

[7] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Ma, J. and Wang, K. (2021) Image preprocessing in classification and identification of diabetic eye diseases. *Data Science and Engineering* **6**(4): 455–471. doi:https://doi.org/10.1007/s41019-021-00167-z.

[8] Sarki, R., Ahmed, K., Wang, H., Zhang, Y. and Wang, K. (2021) Convolutional neural network for multi-class classification of diabetic eye disease. *EAI Endorsed Transactions on Scalable Information Systems* **9**(4). doi:10.4108/eai.16-12-2021.172436.

[9] Cao, Q., Hao, X., Ren, H., Xu, W., Xu, S. and Asiedu, C.J. (2022) Graph attention network based detection of causality for textual emotion-cause pair. *World Wide Web* : 1–15doi:https://doi.org/10.1007/s11280-022-01111-5, URL https://doi.org/10.1007/s11280-022-01111-5.

[10] Xu, S., Song, Y. and Hao, X. (2022) A comparative study of shallow machine learning models and deep learning models for landslide susceptibility assessment based on imbalanced data. *Forests* **13**(11). doi:10.3390/f13111908, URL https://www.mdpi.com/1999-4907/13/11/1908.

[11] Tawhid, M.N.A., Siuly, S., Wang, K. and Wang, H. (2023) Automatic and efficient framework for identifying multiple neurological disorders from eeg signals. *IEEE Transactions on Technology and Society* **4**(1): 76–86. doi:10.1109/TTS.2023.3239526.

[12] Alvi, A.M., Siuly, S. and Wang, H. (2023) A long short-term memory based framework for early detection of mild cognitive impairment from eeg signals. *IEEE Transactions on Emerging Topics in Computational Intelligence* **7**(2): 375–388. doi:10.1109/TETCI.2022.3186180.

[13] Alvi, A.M., Siuly, S., Wang, H., Wang, K. and Whittaker, F. (2022) A deep learning based framework for diagnosis of mild cognitive impairment. *Knowledge-Based Systems* **248**: 108815. doi:https://doi.org/10.1016/j.knosys.2022.108815.

[14] Tawhid, M.N.A., Siuly, S., Wang, H., Whittaker, F., Wang, K. and Zhang, Y. (2021) A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from eeg. *Plos one* **16**(6): e0253094. doi:https://doi.org/10.1371/journal.pone.0253094.

[15] Singh, R., Subramani, S., Du, J., Zhang, Y., Wang, H., Miao, Y. and Ahmed, K. (2023) Antisocial behavior identification from twitter feeds using traditional machine learning algorithms and deep learning. *EAI Endorsed Transactions on Scalable Information Systems* **10**(4). doi:10.4108/eetsis.v10i3.3184.

[16] Pang, X., Ge, Y.F., Wang, K., Traina, A.J. and Wang, H. (2023) Patient assignment optimization in cloud healthcare systems: a distributed genetic algorithm. *Health Information Science and Systems* **11**(1): 30. doi:https://doi.org/10.1007/s13755-023-00230-1.

[17] Pandey, K. and Pandey, D. (2023) Mental health evaluation and assistance for visually impaired people. *EAI Endorsed Transactions on Scalable Information Systems* **10**(4): e6–e6. doi:10.4108/eetsis.vi.2931.

[18] Zhong, Z., Sun, L., Subramani, S., Peng, D. and Wang, Y. (2023) Time series classification for portable medical devices. *EAI Endorsed Transactions on Scalable Information Systems* **10**(4): e19–e19. doi:DOI:10.4108/eetsis.v10i3.3219.

[19] Pandey, D., Wang, H., Yin, X., Wang, K., Zhang, Y. and Shen, J. (2022) Automatic breast lesion segmentation in phase preserved dce-mris. *Health Information Science and Systems* **10**(1): 9. doi:https://doi.org/10.1007/s13755-022-00176-w.

[20] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 3431–3440. doi:https://doi.org/10.1109/CVPR.2015.7298965.

[21] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (Springer): 234–241. URL https://arxiv.org/pdf/1505.04597.pdf.

[22] Tong, G., Li, Y., Chen, H., Zhang, Q. and Jiang, H. (2018) Improved u-net network for pulmonary nodules segmentation. *Optik* **174**: 460–469. doi:https://doi.org/10.1016/j.ijleo.2018.08.086, URL https://doi.org/10.1016/j.ijleo.2018.08.086.

[23] Maji, D., Sigedar, P. and Singh, M. (2022) Attention res-unet with guided decoder for semantic segmentation of brain tumors. *Biomedical Signal Processing and Control* **71**: 103077. doi:https://doi.org/10.1016/j.bspc.2021.103077, URL https://doi.org/10.1016/j.bspc.2021.103077.

[24] Cao, G., Wang, Y., Zhu, X., Li, M., Wang, X. and Chen, Y. (2020) Segmentation of intracerebral hemorrhage based on improved u-net. In *2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*: 183–185. doi:https://doi.org/10.1109/TOCS50858.2020.9339707.

[25] Wen, X., Zhao, B., Yuan, M., Li, J., Sun, M., Ma, L., Sun, C. *et al.* (2022) Application of multi-scale fusion attention u-net to segment the thyroid gland on localized computed tomography images for radiotherapy. *Frontiers in Oncology* **12**: 844052. doi:https://doi.org/10.3389/fonc.2022.844052.

[26] Yu, S., Wang, K., He, L. *et al.* (2022) Pneumothorax segmentation method based on improved u-net network. *Computer Engineering and Applications* **58**(3): 207–214. doi:https://doi.org/10.3778/j.issn.1002-8331.2008-0214.

[27] Cai, S., Xiao, Y. and Wang, Y. (2024) Two-dimensional medical image segmentation based on u-shaped structure. *International Journal of Imaging Systems and Technology* **34**(1): e23023.

[28] Wu, H., Zhang, Z., Zhang, Y., Sun, B. and Zhang, X. (2024) Acx-unet: a multi-scale lung parenchyma segmentation study with improved fusion of skip connection and circular cross-features extraction. *Signal, Image and Video Processing* **18**(1): 525–533.

[29] Kumar, D.N. and Joseph, M.K. (2024) Fused feature vector and dual fcm for lung segmentation from chest x-ray images. *International Journal of Intelligent Engineering & Systems* **17**(2).

[30] Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G. and Zhang, D. (2022) Ds-transunet: Dual swin transformer u-net for medical image segmentation. *IEEE Transactions on Instrumentation and Measurement* **71**: 1–15. doi:https://doi.org/10.1109/TIM.2022.3178991, URL https://doi.org/10.1109/TIM.2022.3178991.

[31] Zhang, J., Du, J., Liu, H., Hou, X., Zhao, Y. and Ding, M. (2019) Lu-net: An improved u-net for ventricular segmentation. *IEEE Access* **7**: 92539–92546. doi:https://doi.org/10.1109/ACCESS.2019.2925060, URL https://doi.org/10.1109/ACCESS.2019.2925060.

[32] Chen, Y., Dai, X., Chen, D., Liu, M., Dong, X., Yuan, L. and Liu, Z. (2022) Mobile-former: Bridging mobilenet and transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*: 5270–5279. doi:https://doi.org/10.48550/arXiv.2108.05895.

[33] Khan, Z.Y. and Niu, Z. (2021) Cnn with depthwise separable convolutions and combined kernels for rating prediction. *Expert Systems with Applications* **170**: 114528. doi:https://doi.org/10.1016/j.eswa.2020.114528, URL https://doi.org/10.1016/j.eswa.2020.114528.

[34] Quiñonez, Y., Lizarraga, C., Peraza, J. and Zatarain, O. (2020) Image recognition in uav videos using convolutional neural networks. *IET Software* **14**(2): 176–181. doi:https://doi.org/10.1049/iet-sen.2019.0045, URL https://doi.org/10.1049/iet-sen.2019.0045.

[35] Zhuxi, M., Li, Y., Huang, M., Huang, Q., Cheng, J. and Tang, S. (2022) A lightweight detector based on attention mechanism for aluminum strip surface defect detection. *Computers in Industry* **136**: 103585. doi:https://doi.org/10.1016/j.compind.2021.103585, URL https://doi.org/10.1016/j.compind.2021.103585.

[36] Fu, H., Song, G. and Wang, Y. (2021) Improved yolov4 marine target detection combined with cbam. *Symmetry* **13**(4): 623. doi:https://doi.org/10.3390/sym13040623, URL https://doi.org/10.3390/sym13040623.

[37] Pan, X., Ge, C., Lu, R., Song, S., Chen, G., Huang, Z. and Huang, G. (2022) On the integration of self-attention and convolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*: 815–825. doi:https://doi.org/10.48550/arXiv.2111.14556.

[38] Canayaz, M. (2021) C+ effxnet: A novel hybrid approach for covid-19 diagnosis on ct images based on cbam and efficientnet. *Chaos, Solitons & Fractals* **151**: 111310. doi:https://doi.org/10.1016/j.chaos.2021.111310.

[39] Taud, H. and Mas, J. (2018) Multilayer perceptron (mlp). *Geomatic approaches for modeling land change scenarios* : 451–455doi:https://doi.org/10.1007/978-3-319-60801-3_27.

[40] van Ginneken, B. and Jacobs, C. (2019), Luna16 part 1/2. URL https://zenodo.org/record/3723295.

[41] van Ginneken, B. and Jacobs, C. (2019), Luna16 part 2/2. URL https://zenodo.org/record/4121926.

[42] Nwankpa, C., Ijomah, W., Gachagan, A. and Marshall, S. (2018) Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378* doi:https://doi.org/10.48550/arXiv.1811.03378, URL https://doi.org/10.48550/arXiv.1811.03378.

[43] Soomro, T.A., Afifi, A.J., Gao, J., Hellwich, O., Paul, M. and Zheng, L. (2018) Strided u-net model: Retinal vessels segmentation using dice loss. In *2018 Digital Image Computing: Techniques and Applications (DICTA)* (IEEE): 1–8. doi:https://doi.org/10.1109/DICTA.2018.8615770.

[44] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(12): 2481–2495. doi:https://doi.org/10.1109/TPAMI.2016.2644615.

[45] Xiao, X., Lian, S., Luo, Z. and Li, S. (2018) Weighted res-unet for high-quality retina vessel segmentation.

In *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*: 327–331. doi:https://doi.org/10.1109/ITME.2018.00080.

[46] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y. [eds.] *Computer Vision – ECCV 2018* (Cham: Springer International Publishing): 833–851. URL https://github.com/tensorflow/models/tree/master/research/deeplab.

[47] Murugappan, M., Bourisly, A.K., Prakash, N., Sumithra, M. and Acharya, U.R. (2023) Automated semantic lung segmentation in chest ct images using deep neural network. *Neural Computing and Applications* : 15343–15364doi:https://doi.org/10.1007/s00521-023-08407-1.

[48] Du, G., Cao, X., Liang, J., Chen, X. and Zhan, Y. (2020) Medical image segmentation based on u-net: A review. *Journal of Imaging Science and Technology* URL https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508.

[49] Jiang, Y., Ye, M., Wang, P., Huang, D. and Lu, X. (2022) Mrf-iunet: A multiresolution fusion brain tumor segmentation network based on improved inception u-net. *Computational and Mathematical Methods in Medicine* **2022**. doi:https://doi.org/10.1155/2022/6305748, URL https://doi.org/10.1155/2022/6305748.