Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/67/20240673

Lung X-ray image segmentation algorithm based on Multihead Self-Attention Mechanism (MSAG) optimizing Unet networks

Haotian Chen

College of Mathematics and Physics, Southwest University of Science and Technology, Sichuan, Mianyang, 621010, China

2665755513@qq.com

Abstract. In this paper, the Multihead Self-Attention Mechanism (MSAG) is used to optimize the Unet network for accurate segmentation of lung X-ray images. By introducing the MSAG module, the ability of the Unet network to capture global and local correlations is enhanced, which effectively improves the accuracy of the segmentation results. The introduction of the multi-head self-attention mechanism enables the network to have more powerful modelling and generalization capabilities, and can process various types of lung X-ray images stably and efficiently. The dataset is divided into training, validation and test sets according to the ratio of 4:3:3. The loss gradually converges during the training process, and the model gradually learns the data features and patterns, and the gap between them and the real labels is gradually reduced. The performance on the validation set is good and no over-fitting occurs, demonstrating the ability to generalize on unseen data. The evaluation metrics on the test set show an IoU of 0.85, a Dice of 0.92, and an Accuracy of 0.88, proving that the model can accurately extract lung features for segmentation. This study has achieved satisfactory results in the field of medical images by optimizing the network structure and introducing new techniques, which are of positive significance for improving the accuracy and efficiency of lung X-ray image segmentation.

Keywords: Multihead Self-Attention Mechanism, Unet Networks, IoU, Image Segmentation.

1. Introduction

Traditional lung X-ray image segmentation methods are usually based on image processing techniques and mathematical models [1]. First, deep learning algorithms can automatically learn feature representations from a large amount of data without the need to manually design feature extractors [2]. Second, deep learning models have strong nonlinear fitting ability, which can better adapt to the complex and variable features of lung X-ray images. In addition, deep learning algorithms are able to achieve end-to-end mapping from the original inputs to the final prediction results through end-to-end training, which simplifies the whole segmentation process.

In the lung X-ray image segmentation task, commonly used deep learning models include U-Net [3], DeepLab [4], Mask R-CNN [5], and so on. These models achieve accurate localization and segmentation of targets such as lung tissues and lesion regions by means of structures such as convolutional neural networks (CNNs). Meanwhile, for different types of lung X-ray image datasets and specific task

DOI: 10.54254/2755-2721/67/20240673

requirements, researchers also improve and optimize these models to increase segmentation accuracy and robustness.

In conclusion, deep learning algorithms are playing an increasingly important role as a powerful tool in the field of lung X-ray image segmentation and will continue to drive the development and innovation in this field. In this paper, lung X-ray image segmentation algorithm based on Multihead Self-Attention Mechanism (MSAG) optimized Unet network is used to segment lung images.

2. Related Work

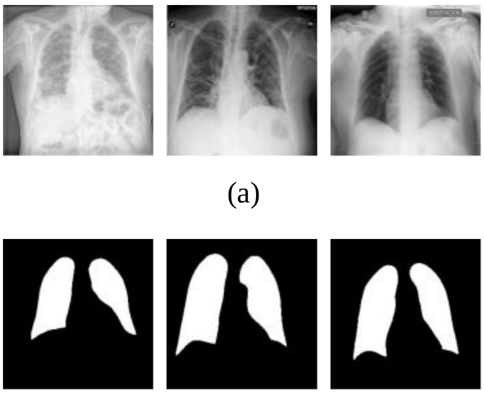
Lung X-ray image segmentation is an important research direction in the field of medical image processing, aiming to accurately segment lung tissues from other tissues (e.g., heart, bones, etc.) in lung X-ray images to help doctors better diagnose and treat patients. With the development of deep learning techniques, significant progress has been made in lung X-ray image segmentation, improving the accuracy and stability of segmentation results [6,7].

In the past few years, many researchers have worked on work related to lung X-ray image segmentation. The commonly used methods in lung X-ray image segmentation include traditional methods based on threshold, edge detection, and region growing, as well as methods based on deep learning. Among them, deep learning methods such as U-Net and Mask R-CNN are widely used in lung X-ray image segmentation tasks and have achieved better results [8].

These methods are trained on a large number of datasets by constructing an end-to-end neural network model and use techniques such as convolutional neural networks to achieve accurate segmentation of different tissue structures in lung X-ray images. These methods not only improve the accuracy and robustness of the segmentation results, but also can effectively reduce the cost and time consumption of manual intervention.

3. Source of Data Sets

The data used in this paper is selected from the Kaggle public dataset, which is available at https://www.kaggle.com/datasets/newra008/lung-mask-image-dataset?rvi=1. The lung X-ray image dataset contains the original lung X-ray images and the corresponding segmentation masks that used to mark out the lung regions in the image. The original lung X-ray images are stored in grey scale image format and the corresponding masks are represented as binarised images, where the lung regions are marked in white and other tissues in black. Some of the lung X-ray images and their masks are shown in Figure 1.



(a)

(b)

Figure 1. Partial image data. (a) Original figure. (b) Mask. (Photo credit: Kaggle public dataset)

DOI: 10.54254/2755-2721/67/20240673

4. Method

Lung X-ray image segmentation algorithm based on Multihead Self-Attention Mechanism (MSAG) optimised U-Net network is an approach that combines deep learning and self-attention mechanism, aiming to improve the accuracy and efficiency of lung X-ray image segmentation task.

4.1. U-Net network

U-Net is a deep learning network structure commonly used for medical image segmentation tasks, which is characterised by a symmetric U-shaped structure, including two parts: an Encoder and a Decoder. The schematic diagram of the Unet network is shown in Figure 2.

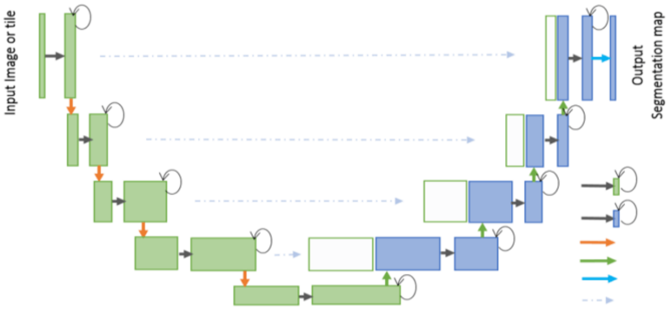


Figure 2. The schematic diagram of the Unet network. (Photo credit: Original)

4.2. Multiple Self-Attention Mechanism (MSAG)

Multi-Head Self-Attention Mechanism (MHSAM) is an important technique used in the field of deep learning for processing sequence data, and has achieved remarkable success especially in natural language processing tasks. The mechanism enables the model to simultaneously focus on information at different locations in the input sequence by introducing multiple attention heads, thus better capturing the dependencies and important features between sequences. The principle of the multi-head self-attention mechanism is described in detail below.

In a multi-head self-attention mechanism, multiple parallel attention heads are usually used to process the input sequence. Each attention head learns different semantic information and computes the attention weights with independently learnt parameters, thus improving the model's ability to characterise the sequence information. Finally, the outputs of multiple heads are spliced together and linearly transformed to obtain the final output representation.

Overall, by introducing multiple parallel computing attention heads and combining the interaction between query, key and value, the multi-head self-attention mechanism enables the model to better capture the complex dependencies among the input sequences and effectively extract important features. This mechanism has not only achieved remarkable results in natural language processing tasks, but also has a wide range of application prospects in other fields such as computer vision [9]. A schematic diagram of the Multi-head Self-Attention Mechanism (MSAG) is shown in Figure 3.

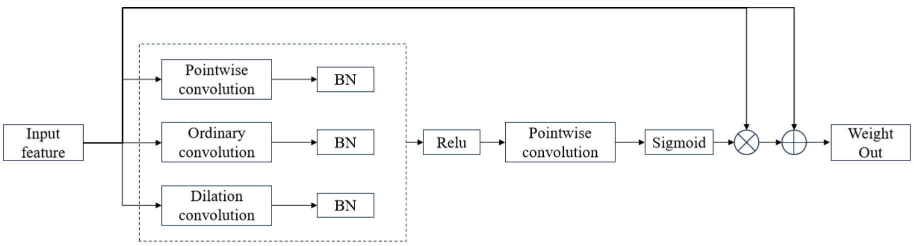


Figure 3. A schematic diagram of the Multi-head Self-Attention Mechanism (MSAG). (Photo credit: Original)

DOI: 10.54254/2755-2721/67/20240673

4.3. MSAG optimises U-Net network

In the lung X-ray image segmentation task, the traditional U-Net network may suffer from insufficient dependence on distant pixels and insufficient use of local information. Optimising the U-Net network by introducing the MSAG module can enable the network to better capture the correlation between global and local, thus improving the accuracy of the segmentation results [10].

Specifically, with the inclusion of the MSAG module in the U-Net network, the multi-head self-attention mechanism can be used in both the encoder and decoder stages to learn the correlations between different layers as well as between different spatial locations [11]. The encoder stage uses the self-attention mechanism to better capture global information and to help the network better understand the overall contextual environment, while the decoder stage can be guided by the self-attention mechanism to generate detailed information more accurately and avoid excessive smoothing or distortion [12].

5. Algorithm Flow and Algorithm Advantages

5.1. Algorithmic process

Encoder: Input raw lung X-ray image is convolved operation to get high level feature representation.

MSAG module: MSAG module is introduced at each stage of encoder and decoder to achieve global and local correlation learning.

Decoder: features are gradually reduced to their original size using up-sampling operations and fused with the corresponding layer encoder features.

Output layer: ultimately generates a prediction mask by convolution operation and computes the loss function with the real mask for training.

5.2. Dominance

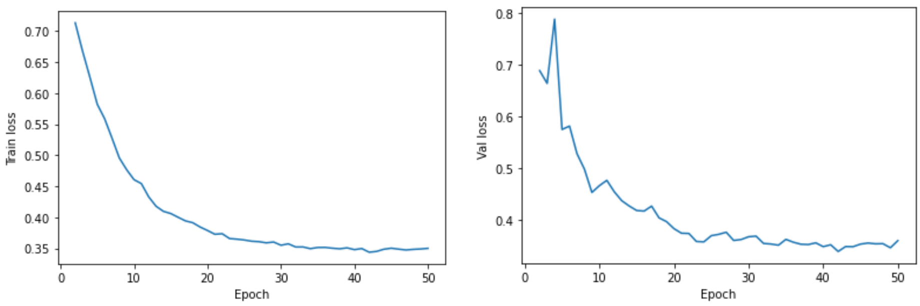
The lung X-ray image segmentation algorithm based on MSAG-optimised U-Net network is able to better combine global and local information, and improve the recognition ability of detailed information such as lung region boundaries in complex scenes while maintaining the simplicity and effectiveness of the U-Net structure. By introducing the multi-head self-attention mechanism, the network is equipped with more powerful modelling and generalisation capabilities, and performs stably and accurately in processing various types of lung X-ray images.

Overall, the lung X-ray image segmentation algorithm based on MSAG-optimised U-Net network combines deep learning and the self-attention mechanism, which has a wide range of prospects for application in the field of medical image processing and brings important help to improve the diagnostic efficiency of doctors and the quality of patient treatment.

6. Result

Firstly the dataset is divided into training set, validation set and test set in the ratio of 4:3:3, 40% of the data is used for training, 30% of the data is used for validation and finally 30% of the data is used for testing. The size of the input image is set to 512x512, the number of initial training rounds is set to 0, the number of frozen training rounds is set to 50, the batch size is set to 2, the initial learning rate is set to 1e-4, and the minimum learning rate is 0.01 times the initial learning rate.

Output the loss change of training set and validation set during the training process, and the results are shown in Figure 4.



(a) (b)

Figure 4. Output the loss change of training set and validation set. (a) Train loss. (b) Val loss. (Photo credit: Original)

The loss in the training set decreases from 0.70 to 0.35 and tends to converge, the model gradually learns the features and patterns of the data during the training process, and the gap between the prediction results and the real labels on the training set decreases.

The loss in the validation set decreases from 0.80 to 0.35 and tends to converge, the model also performs well on the unseen validation dataset without overfitting, and performs well on the training set has good generalisation ability on the validation set.

The model is tested using the test set, the first column is the original image, the second column is the mask of the image and the third column is the predicted output of the test set and the results are shown in Figure 5.

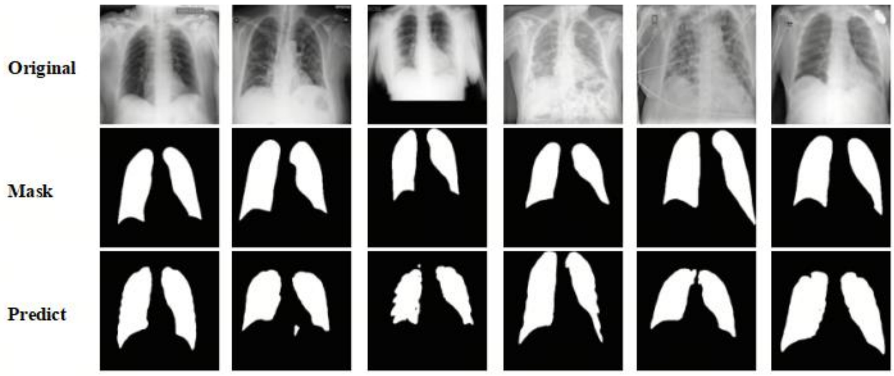


Figure 5. Test Set Output Results. (Photo credit: Original)

The evaluation metrics of the output test set, as shown in Table 1, are 0.85 for IoU, 0.92 for Dice, and 0.88 for Accuracy, and the model is able to segment the lung X-images well and extract the lung features.

Table 1. Modelling Evaluation.

Evaluation Parameters	**Result**
IoU	0.85
Dice	0.92
Accuracy	0.88
F1 Score	0.83
Precision	0.9
Recall	0.85

7. Conclusion

The aim of this study is to improve the accuracy of lung X-ray image segmentation by optimising the U-Net network using the Multihead Self-Attention Mechanism (MSAG). The traditional U-Net network may suffer from insufficient local information capture and weak global correlation when processing medical images, while the introduction of the MSAG module is able to better capture the correlation between global and local, which can effectively improve the segmentation results. The introduction of the multi-head self-attention mechanism makes the network have more powerful modelling and generalization capabilities, and performs stably and accurately in processing various types of lung X-ray images. This paper aims to enhance the precision of lung X-ray image segmentation by optimizing the U-Net network through the incorporation of the Multihead Self-Attention Mechanism (MSAG). The conventional U-Net network may be susceptible to deficiencies in the capture of local information and a lack of robust global correlation when processing medical images. However, the integration of the MSAG module enables more comprehensive correlation between global and local information, thereby facilitating more accurate segmentation outcomes. The incorporation of the multi-head self-attention mechanism imbues the network with enhanced modelling and generalization capabilities, enabling it to process a diverse range of lung X-ray images with stability and accuracy.

To ascertain the efficacy of the optimized U-Net network in the lung X-ray image segmentation task, the dataset was divided into training sets, validation sets, and test sets under a ratio of 4:3:3, as detailed in this paper. During the training process, it was observed that the loss of the training set decreased from 0.70 to 0.35 and converged, indicating that the model gradually learned the data features and patterns and improved the prediction accuracy. Similarly, the loss of the validation set also decreased from 0.80 to 0.35 and converged, verifying the model's ability to generalize to unseen data. Finally, the performance of the MSAG-optimized U-Net network was evaluated on the test set, and the results showed that the Intersection over Union (IoU) was 0.85, the Dice coefficient was 0.92, and the accuracy (Accuracy) was 0.88. These metrics indicate that the MSAG-optimized U-Net network proposed in this study performs well in the lung image segmentation task performs well and has high application potential and accuracy. The objective of this research is to demonstrate the value of integrating global and local information and utilising the multi-head self-attention mechanism to enhance the performance of a deep learning model. It is believed that the results of this research will contribute to the advancement of medical imaging and lay the foundation for further optimization of deep learning models to play a greater role in medical image analysis in the future.In order to verify the effectiveness of the optimized U-Net network in the lung X-ray image segmentation task, we divided the dataset into training, validation and test sets in the ratio of 4:3:3. During the training process, it is observed that the loss of the training set decreases from 0.70 to 0.35 and converges, which indicates that the model gradually learns the data features and patterns and improves the prediction accuracy; while the loss of the validation set also decreases from 0.80 to 0.35 and converges, which verifies the model's good ability to generalize to unseen data.

Finally, the performance of the MSAG-optimized U-Net network was evaluated on the test set, and the results showed that the Intersection over Union (IoU) was 0.85, the Dice coefficient was 0.92, and the accuracy (Accuracy) was 0.88.These metrics indicate that the MSAG-optimized U-Net network proposed in this study performs well in the lung image segmentation task performs well and has high application potential and accuracy. By better fusing global and local information and effectively using the multi-head self-attention mechanism to improve the model performance, we believe that the results of this research will bring important value to the field of medical imaging and lay the foundation for further optimization of deep learning models to play a greater role in medical image analysis in the future.

References

[1] Minki K and Lee B D. “Automatic lung segmentation on chest X-rays using self-attention deep neural network.” Sensors 21.2 (2021): 369.

[2] Arvind S., et al. "Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest X-rays." Results in Engineering 17 (2023): 100929.

[3] Candemir, Sema, and Sameer Antani. "A review on lung boundary detection in chest X-rays." International journal of computer assisted radiology and surgery 14 (2019): 563-576.

[4] Selvan, Raghavendra, et al. "Lung segmentation from chest X-rays using variational data imputation." arxiv preprint arxiv:2005.10052 (2020).

[5] Rashid, Rabia, Muhammad Usman Akram, and Taimur Hassan. "Fully convolutional neural network for lungs segmentation from chest X-rays." Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15. Springer International Publishing, 2018.

[6] Tang, You-Bao, et al. "Xlsor: A robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation." International Conference on Medical Imaging with Deep Learning. PMLR, 2019.

[7] Wang, Hongyu, et al. "A multi-objective segmentation method for chest X-rays based on collaborative learning from multiple partially annotated datasets." Information Fusion 102 (2024): 102016.

[8] Santosh, K. C., et al. "Edge map analysis in chest X-rays for automatic pulmonary abnormality screening." International journal of computer assisted radiology and surgery 11 (2016): 1637-1646.

[9] Ye, Run Zhou, et al. "Automatic ARDS surveillance with chest X-ray recognition using convolutional neural networks." Journal of Critical Care 82 (2024): 154794.

[10] Goshima, Futa, et al. "Deep learning-based algorithm to segment pediatric and adult lungs from dynamic chest radiography images using virtual patient datasets." Medical Imaging 2024: Physics of Medical Imaging. Vol. 12925. SPIE, 2024.

[11] Samip G and Subedi S. "Estimating Lung Volume Capacity from X-ray Images Using Deep Learning." Quantum Beam Science 8.2 (2024): 11.

[12] Killeen, Benjamin D., et al. "FluoroSAM: A Language-aligned Foundation Model for X-ray Image Segmentation." arxiv preprint arxiv:2403.08059 (2024).