BILINGUAL PUBLISHING GROUP

Journal of Computer Science Research

https://journals.bilpubgroup.com/index.php/jcsr

ARTICLE

Improving U-Net Performance for Tumor Segmentation Using Attention Mechanisms

Zetai Wu, Weifa Liu, Jing Chang*

Guangzhou City University of Technology, Guangzhou 510800, China

ABSTRACT

U-Net is a widely recognized neural network model for medical image segmentation, renowned for its efficiency in extracting features from both current and past input data. However, traditional U-Net models exhibit limitations in extracting edge features, particularly in medical CT images characterized by complex gray distributions and close pixel intervals. This leads to suboptimal performance, with low accuracy, recall, intersection over union (IoU), and F1-score. This research proposes an improved U-Net model incorporating an attention mechanism to enhance tumor segmentation accuracy and efficiency. The attention mechanism strategically weights important features, directing the network to focus on task-relevant areas. Experimental results demonstrate that our proposed attention-based U-Net model significantly improves tumor segmentation performance, achieving notable enhancements in accuracy, recall, IoU, and F1-score. Further validation across diverse datasets confirms the model's generalization ability and superiority compared to the original U-Net method. This research contributes to the advancement of medical image segmentation techniques, highlighting the potential of attention mechanisms in optimizing deep learning models for clinical applications.

Keywords: Deep learning; Medical image segmentation; Unet model; Attention mechanism

*CORRESPONDING AUTHOR:

Jing Chang, Guangzhou City University of Technology, Guangzhou 510800, China; Email: changjing@gcu.edu.cn

ARTICLE INFO

Received: 12 September 2024 | Revised: 20 September 2024 | Accepted: 25 September 2024 | Published Online: 27 September 2024

DOI: https://doi.org/10.30564/jcsr.v6i4.7271

CITATION

Wu, Z., Liu, W., Chang, J., 2024. Improving U-Net Performance for Tumor Segmentation Using Attention Mechanisms. Journal of Computer Science Research. 6(4): 66–72. DOI: https://doi.org/10.30564/jcsr.v6i4.7271

COPYRIGHT

Copyright © 2024 by the author(s). Published by Bilingual Publishing Co. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (https://creativecommons.org/licenses/by-nc/4.0/).

1. Introduction

The field of medical image segmentation has always been one of the research directions in the field of medical image processing, and it has important application value in assisting clinical diagnosis and treatment planning [1]. In recent years, with the rapid development of deep learning technology, image segmentation methods based on deep learning have gradually become a research hotspot. As a classic deep learning network structure, Unet has been widely used in medical image segmentation tasks [2]. However, although Unet performs well in dealing with general image segmentation tasks, its performance still has some shortcomings in complex scenes. Aiming at the problems existing in Unet when dealing with complex scenes, this study introduces an attention mechanism, aiming to make the network pay more attention to task-related regions by weighting important features, thereby improving the accuracy and robustness of image segmentation. The introduction of attention mechanism enables the network to adjust the degree of attention to different areas more flexibly, so as to better adapt to the segmentation tasks in different scenarios. The experimental results show that our proposed Unet model based on improved attention mechanism has made significant improvements in tumor segmentation tasks. This study will first introduce the basic principle of data and preprocessing method and Unet model and its application in medical image segmentation. Then, the improved method of introducing attention mechanism will be described in detail, and its effectiveness will be proved by experiments. Finally, we will discuss the limitations and future improvement directions of this method, as well as its potential application value in the field of medical image segmentation.

2. Related work

2.1 Windowing method

Windowing is a method in medical image processing, which is used to adjust the gray components of CT images to highlight specific structures. Because the HU value range of medical imaging is generally very large, it will generally lead to poor contrast. If you train with this kind of picture, you will get poor results. So we need to use windowing method to process the data first. The effect is shown in **Figure 1**.

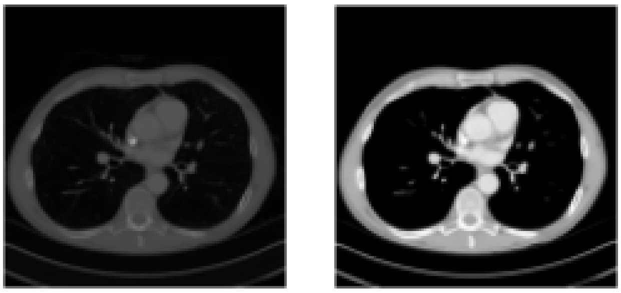

Figure 1. Comparison of CT images before and after using windowing method.

First, an image in the DICOM format is converted to an HU value (CT image value), which is not required if the image is already a CT image. By adjusting the window width and window level, we can improve the contrast of the image and highlight the structure of interest. The windowing adjustment formula is as follows:

$$f\left(x\right)=\left\{\begin{array}{c} 0,\frac{x-winCenter+\frac{winWidt}{2}}{winWidt}<0 \\ 1,\frac{x-winCenter+\frac{winWidt}{2}}{winWidt}>1 \\ x,0<\frac{x-winCenter+\frac{winWidt}{2}}{winWidt}<1 \end{array}\right\} \tag{1}$$

where winWidth represents the width of window, winCenter represents the center of window.

2.2 Histogram equalization

If we need to improve the contrast and visual effect of the image, then we need to perform histogram equalization on the image. It is a method commonly used for medical image preprocessing. Usually, we need to make the gray level distribution of the image more uniform, so as to improve the contrast and make the image clearer. Moreover, we can redistribute the pixel values of the image to expand those pixel values that occupy a smaller gray range in the original image. Histogram equalization can improve the contrast of CT images, improve image quality, and provide doctors with more accurate diagnostic information [3] in medical images. For muscles and soft tissues in CT images, histogram equalization can make the gray distribution more uniform and improve the recognition effect. The histogram equalization formula can be defined as Equation (2).

$$F\left(x\right)=\sum\nolimits_{i=0}^{k}\frac{n_i}{n},k=0,1,2,...,L-1 \tag{2}$$

N is the sum of the number of pixels in the image, ni is the number of pixels in the current gray level, and L is the total number of gray levels of the image. The effect is shown in **Figure 2**.

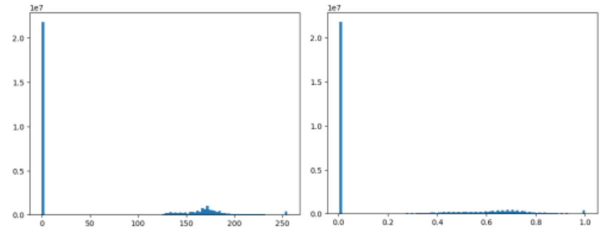

Figure 2. Distribution of pixel data after histogram equalization. This method has a wide range of applications in medical image processing and natural scene image enhancement, which can effectively improve image quality and the accuracy of image analysis and diagnosis.

2.3 Biclassification cross entropy

The loss function of the model is 0-1 binary_crossentropy, and its characteristic is that if the predicted value is wrong, the corresponding value of the loss function will be very large. The 0-1 binary classification cross entropy can be defined as Equation (3),

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p\,(\hat{y}_i)) + (1-y_i)\cdot log(1-p\,(\hat{y}_i)) \quad (3)$$

Another reason for choosing this loss function is that UNet network is usually used for segmentation tasks, and the output is multi-channel. The cross-entropy loss function can be directly applied to multi-channel prediction without additional modification.

3. Methodology

3.1 Attention module design

Attention mechanism is a module that enhances the ability to extract image features in model, improves the model's attention to feature parts, and weights important information[4]. In this paper, self-attention mechanism is used to strengthen the discrimination degree of feature maps and reduce unnecessary redundant information. Due to the self-attention mechanism, the corresponding convolution layer need to be formed to convolve it and store it as two matrices for the processed feature maps.

The specific operation steps are as follows:

Step 1: Perform a convolution operation on the input signal, and use a sequential model including a convolution layer and a batch normalization layer to map the input signal to a new feature map to obtain feature maps F1 and F2. Where the filter size of the convolutional layer is 1 and the step size is 1.

Step 2: The feature matrices corresponding to the feature maps F1 and F2 are spliced and activated by using and using the Relu function.

Step 3: Reduce the channel to 1, continue to take convolution operations on the merged feature maps, and regularize them. For each batch of data, the output at each layer of the network is normalized, that is, the value of each feature is scaled to a standard normal distribution with a mean value of 0 and a standard deviation of 1.

Finally, Sigmoid function is applied to obtain the weight matrix, and finally the weight matrix is returned.

3.2 Unet + attention network model

Adding Attention mechanism to Unet network can enhance the detection ability of small target characteristics and improve the perception range[5]. The direct embodiment is to improve the accuracy and intersection ratio of the model. The structure of the model is shown in **Figure 3**. The network model consists of input layer, output layer, three encoders and three decoders. An encoder consists of two convolutional layers and one pooling layer. A decoder consists of two convolutional layers and one deconvolutional layer. The data output by each encoder will not only be used as the input of the next layer, but also as a part of the input of the corresponding decoder. That is, the input source of the decoder is the connection composition of the output of the decoder of the upper layer and the output of the corresponding encoder.

4. Experiments

4.1 Experimental setup

The python library used is as follows.

NumPy is an open-source library for Python that provides powerful tools for numerical computation and array manipulation. It offers a highly efficient way to store and process large arrays, significantly outperforming Python's built-in nested lists for such tasks. NumPy supports multi-dimensional arrays, enabling sophisticated matrix operations. Moreover, it comes with a comprehensive set of mathematical functions specifically designed for array operations, simplifying and accelerating scientific computing workflows.

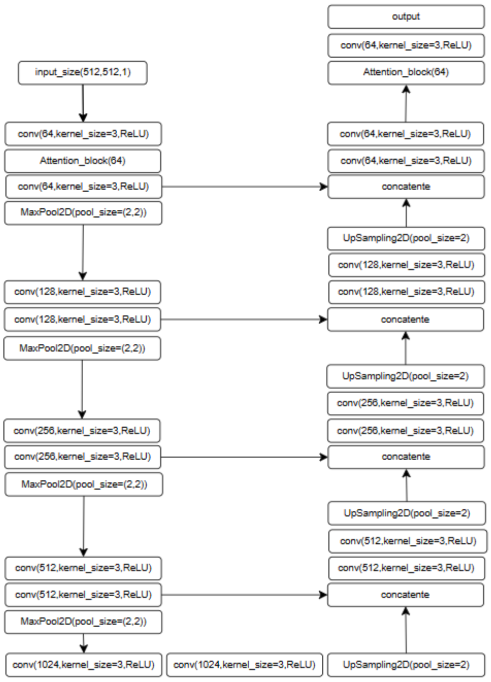
input_size(512,512,1)
conv(64,kernel_size=3,ReLU)
Attention_block(64)
conv(64,kernel_size=3,ReLU)
MaxPool2D(pool_size=(2,2))
conv(128,kernel_size=3,ReLU)
conv(128,kernel_size=3,ReLU)
MaxPool2D(pool_size=(2,2))
conv(256,kernel_size=3,ReLU)
conv(256,kernel_size=3,ReLU)
MaxPool2D(pool_size=(2,2))
conv(512,kernel_size=3,ReLU)
conv(512,kernel_size=3,ReLU)
MaxPool2D(pool_size=(2,2))
conv(1024,kernel_size=3,ReLU)
conv(1024,kernel_size=3,ReLU)
UpSampling2D(pool_size=2)
concatente
conv(512,kernel_size=3,ReLU)
conv(512,kernel_size=3,ReLU)
UpSampling2D(pool_size=2)
concatente
conv(256,kernel_size=3,ReLU)
conv(256,kernel_size=3,ReLU)
UpSampling2D(pool_size=2)
concatente
conv(128,kernel_size=3,ReLU)
conv(128,kernel_size=3,ReLU)
UpSampling2D(pool_size=2)
concatente
conv(64,kernel_size=3,ReLU)
conv(64,kernel_size=3,ReLU)
Attention_block(64)
conv(64,kernel_size=3,ReLU)
output

Figure 3. Unet+Attention model structure diagram.

PyTorch is the Python-based version of Torch, a popular open-source deep learning framework originally developed by Facebook. It's specifically designed for building and training deep neural networks (DNNs), leveraging the power of GPUs for accelerated computation. Torch, the underlying foundation of PyTorch, is a robust tensor library that handles multidimensional array data, making it a versatile tool for machine learning and other computationally demanding tasks.

The software and hardware configuration of the experimental environment is shown in the **Table 1**.

Table 1. The software and hardware configuration.

CPU	**Graphics card**	**Memory**
vCPU Intel(R) Xeon(R)	L20	48G
Operating system	**Python**	**CUDA**
Ubuntu20.04	3.8	11.2
Tensorflow	**IDE**	**Platform**
2.9.0	Pycharm2023	AutoDL

4.2 Data sources

We use two kinds of dataset from open source Library and use the first 70% as the training set and the last 30% as the test set.

(1) Dataset Name: MSD Lung Cancer Segmentation. Source statement of lung tumor CT data: The research data comes from the institutional release–Cornell University (Cornell University);

(2) Dataset name: 3Dircadb1. Liver Tumor CT Data Source Statement: Research data from institutional release: French Institute of Gastrointestinal Cancer;

4.3 Evaluation Metrics

To evaluate the accuracy of image predictions, we employ five metrics: Accuracy, Precision, Recall, F1 Score, and Intersection over Union (IOU). Prior to calculating these metrics, we must generate a confusion matrix based on the pixel-level predictions of the image, as illustrated in **Table 2**.

Table 2. Confusion matrix.

Confusion		**Predicted value**	
		Positive	**Negative**
Actual Value	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)

(1) Accuracy

Accuracy is the most straightforward evaluation indicator, which is the ratio of the number of samples correctly classified by the model to the total number of samples. However, it may not be enough to accurately evaluate the model performance in the case of unbalanced category distribution [6].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

(2) Precision

Precision measures the proportion of correctly identified positive cases among all cases predicted as positive by the model. In other words, it tells us how accurate the model is at predicting positive cases. A high precision score indicates that the model is good at identifying true positives while minimizing false positives [7].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

(3) Recall

Recall, also known as sensitivity or the true positive rate, measures the proportion of actual positive cases that the model correctly identifies as positive. It tells us how well the model is able to find all the positive cases. A high recall score indicates that the model is good at identifying true positives, minimizing false negatives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

(4) F1 Score

The F1 score is the harmonic mean of precision and recall, providing a single metric that reflects the balance between these two performance measures. It represents a weighted average of precision and recall, effectively balancing their importance. A higher F1 score indicates that the model achieves a better balance between correctly identifying positive cases (precision) and finding all actual positive cases (recall).

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

(5) IOU

Intersection over Union (IoU), also known as the Jaccard Index, is a commonly used metric to evaluate tasks such as object detection or semantic segmentation. It measures the overlap between the predicted result (bounding box or segmentation mask) and the actual ground truth result. Specifically, IoU is calculated as the ratio of the area of intersection between the predicted and ground truth regions to the area of their union. A higher IoU score indicates a better alignment between the model's prediction and the actual result.

$$\text{IOU} = \frac{\text{Prediction Images} \cap \text{Real Images}}{\text{Prediction Images} \cup \text{Real Images}} \tag{8}$$

4.4 Experimental design

This experiment utilizes a three-dimensional dataset that undergoes a preprocessing pipeline involving windowing and histogram equalization. The preprocessed data is then converted into image format, with each image having a pixel dimension of 512 x 512 x 1, representing a single channel (e.g., grayscale). Prior to model training, the Adam optimizer is configured with the following hyperparameters:

Learning Rate: 0.0002. This value controls the step size taken during parameter updates during optimization.

The selected learning rate is a common starting point for Adam optimization, providing a balance between convergence speed and avoiding oscillations in the loss function.

Epochs: 50. This indicates the number of complete passes through the entire training dataset

The choice of 50 epochs allows for sufficient training iterations to learn the underlying patterns in the data.

Steps per Epoch: 200. This determines the number of training updates performed within each epoch.

The steps per epoch setting ensures that the model encounters a diverse set of training examples within each epoch, promoting more robust learning.

4.5 Experimental results

We employed PyCharm to connect to AutoDL's remote GPUs for model training, with the trained model results saved to the corresponding path in Jupyter notebook. The test dataset was assembled by combining a subset of samples but were not used in the training process as described in Section 4.2.

We first adopt the traditional Unet network trained as the baseline model. Results demonstrated suboptimal performance, characterized by low accuracy, recall, F1-score, and IoU values. Furthermore, the model exhibited errors in boundary detection and inaccurate positional predictions despite accurate orientation.

Subsequently, we retest the Unet network was enhanced by incorporating the Attention mechanism. The results indicated a significant improvement in the performance of the modified Unet model with Attention. As depicted in **Table 3** and illustrated in **Figure 4**, the inclusion of the Attention mechanism resulted in a substantial increase in accuracy, recall, F1-score, and IoU values, approximately doubling these metrics compared to the original Unet model. These findings confirm the effectiveness of incorporating the Attention mechanism in improving the accuracy of the traditional Unet model[8].

Table 3 presents a comprehensive analysis of the performance metrics on two distinct datasets: MSD (for lung tumor segmentation) and 3Dircadb1 (for liver tumor segmentation). The results highlight the robustness of the enhanced Unet model with Attention, demonstrating its superior performance across diverse data sets.

Table 3. Experimental metrics results.

	MSD		3Dircadb1	
	Unet	**AttentionUnet**	**Unet**	**AttentionUnet**
Accuracy	0.99	0.99	0.99	0.99
Precision	0.21	0.78	0.36	0.71
Recall	0.31	0.46	0.36	0.58
F1 Score	0.20	0.56	0.36	0.64
IoU(Jaccard Score)	0.13	0.41	0.23	0.48

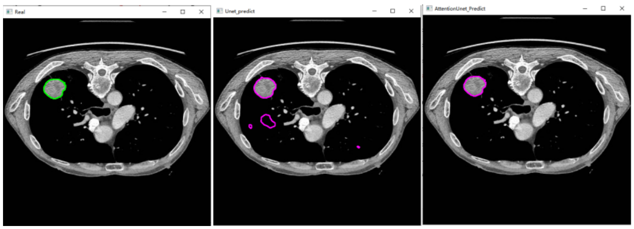

Figure 4. Comparison of MSD dataset real image (left), Unet model prediction map (middle) and AttentionUnet model prediction map (right).

Figure 4 visually compares the segmentation results of the original Unet model and the Unet model with Attention on a representative image from the MSD dataset. This comparison further underscores the improvement in accuracy and localization of the segmented tumor region achieved by the enhanced model.

These findings clearly demonstrate that incorporating an attention mechanism into the conventional Unet model significantly enhances its performance in tumor segmentation tasks. The improved accuracy, recall, F1-score, and IoU values, coupled with the validation on diverse datasets, solidify the effectiveness and generalizability of the enhanced Unet model with Attention [9].

This research contributes to the advancement of medical image segmentation techniques by demonstrating the effectiveness of integrating the attention mechanism within a widely recognized deep learning architecture. Further exploration of this approach could involve investigating the impact of different attention mechanisms and their integration with other deep learning architectures for improved performance in various medical image segmentation tasks.

5. Conclusion

The U-Net model has proven to be a widely validated architecture in the field of medical image segmentation. However, the original U-Net model exhibits certain limitations in practical applications. This research proposes a novel approach to enhance tumor segmentation by implementing a U-Net model augmented with an attention mechanism. This hybrid model, denoted as U-Net+Attention, addresses the shortcomings of the conventional U-Net architecture through the strategic incorporation of a custom-designed attention block.

The experimental results consistently demonstrate a significant improvement in accuracy with the U-Net+Attention model. By analyzing the performance across different datasets, we validate the effectiveness of our approach in enhancing the precision of tumor segmentation. The enhanced model demonstrates a notable improvement in its ability to generalize well across diverse datasets.

The proposed U-Net+Attention model provides a promising avenue for advancing tumor segmentation capabilities in medical imaging. It opens avenues for future research to explore various attention mechanisms and their integration with other deep learning architectures, aiming to achieve further refinements in performance across a wider spectrum of medical image segmentation tasks.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

This work is supported by the Open Foundation of the Guangdong Provincial Science and Technology Innovation Strategy Special Fund (pdjh2023b0784).

References

[1] Zaremba, W., Sutskever, I., Vinyals, O., 2014. Recurrent neural network regularization. arXiv preprint. arXiv:1409.2329.

[2] Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net:

Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., et al. (eds) Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science. Springer: Cham. 9351, 234–241.

[3] Bao, W., Yue, J., Rao, Y., 2017. A deep learning framework for financial time series using stacked autoencoders and longshort term memory. PloS one. 12(7), e0180944.

[4] Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint. arXiv:1803.01271.

[5] Jiang, H., Qiu, X., Chen, J., et al., 2019. Insulator fault detection in aerial images based on ensemble learning with multi-level perception. IEEE Access. 7, 61797–61810.

[6] Tao, X., Zhang, D., Wang, Z., et al., 2020. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 50(4), 1486–1498.

[7] Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing. 452, 48–62.

[8] Sun, Y., Bi, F., Gao, Y., et al., 2022. A multi-attention UNet for semantic segmentation in remote sensing images. Symmetry. 14(5), 906.

[9] Peng, C., Zhang, H., Wang, Y., 2022. Fast Detection Method of Insulator String Image Basedon YOLOv3. Insulators and Surge Arresters. (01), 151–156.