

CCUNet: UNet Based on an Improved Coordinate Channel Attention Mechanism and Its Applications



Ning Ma¹, Songwen Jin^{2*}

¹Tangshan Normal University, Tangshan 063000, China

²College of Art and Design of Hebei Minzu Normal University, Chengde 067000, China

Corresponding Author Email: jsw19810718@sina.com

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420111>

ABSTRACT

Received: 8 July 2024

Revised: 18 December 2024

Accepted: 22 December 2024

Available online: 28 February 2025

Keywords:

deep learning, self-attention mechanism, UNet, Tversky loss function

To address the issues of small lesion omission, misclassification of strongly similar tissues, and data imbalance in UNet-based medical image processing, the paper was proposed. The algorithm improves the coordinate channel attention mechanism by constructing fused attention in three directions-horizontal, vertical, and channel-thereby building a 3D dependency of the image and enhancing the attention mechanism's focus on nodules. Through sub-attention mechanisms, the coordinate channel attention mechanism was further refined to address misclassification in complex backgrounds or regions with similar intensity. Additionally, the Tversky loss function was improved to increase the model's sensitivity to minority classes. Experiments were performed on the LUNA16 dataset, and a comparison with traditional algorithms demonstrated that the proposed algorithm is feasible, advantageous, and stable, effectively addressing the aforementioned issues.

1. INTRODUCTION

In 2020, there were 2.2 million new cases of lung cancer globally, with the mortality rate approaching four-fifths, making lung cancer the leading cause of cancer-related deaths among all cancers [1]. Only one-fifth of patients are diagnosed at the early stage of the pathological process [2, 3], receiving effective treatment that extends their lifespan and reduces mortality. Among the four stages of lung cancer, stage 1 and 2 lung cancers have significantly better treatment outcomes compared to stage 3 or 4 lung cancers [4]. Therefore, the early detection and treatment of asymptomatic lung cancer are effective means to increase lifespan and decrease mortality.

Computed tomography (CT) serves as an advanced non-invasive imaging technique for precise nodule localization, tumor size estimation, morphological assessment, prognosis determination, and tumor classification based on various attributes and life cycle stages [5, 6]. Evidence from the National Lung Cancer Screening Trial (NLST) [7, 8] suggests that the implementation of low-dose computed tomography (LDCT) screening in high-risk groups can deliver high-fidelity pulmonary imaging while minimizing radiation exposure, ensuring robust characterization and dynamic tracking of pulmonary nodules, which contributes to a 20% reduction in lung cancer mortality. However, the complexity of CT image backgrounds, the tiny dimensions of pulmonary nodules, and the inherent data imbalance in CT scans contribute to the difficulty of current computational algorithms in precisely distinguishing between Non-cancerous tissue nodules and cancerous tumor nodules [9].

Thus, optimizing CT image analysis models and refining or developing more precise models for the segmentation,

classification, and prediction of pulmonary nodules is of critical importance.

Deep learning technology [10, 11], which effectively combines convolutional neural networks, has been extensively utilized in medical image processing and has resulted in remarkable achievements [12, 13]. Radiomics is a cutting-edge technology that enables the extraction of high-dimensional, quantitative image features for clinical diagnostic evaluation and prognostic forecasting [14]. Texture features in medical images [15] are crucial for clinical diagnosis and serve as a key foundation for lung cancer diagnosis. In the classification, prediction, and malignant risk assessment of pulmonary nodules, deep convolutional neural networks have achieved notable progress.

However, in medical image analysis, deep convolutional neural networks also have the following problems:

(1) As the depth increases, the high abstraction of features leads to the loss of small targets (such as pulmonary nodules). (2) Convolutional neural networks can only learn local spatial features, ignoring global features and long-distance dependency learning. (3) During the learning process, convolutional neural networks lack the ability to highlight task-relevant features, resulting in the suppression or disregard of unrelated components.

Based on this, the UNet neural network and attention mechanism are introduced in medical image analysis to address the above issues. The U-Net algorithm [16-18] facilitates multi-feature detection in a layer-wise process, addressing the challenge of inability to preserve important characteristics of small targets. By employing global computation methods, it captures the global features of the data, Overcoming the restriction of convolutional neural

networks that are limited to extracting local features. Attention mechanism-based model [19-21] to highlight task-relevant objects or parts and ignore irrelevant ones. The fusion of multiple attention mechanisms is used to address long-distance dependencies, enhancing the model's learning capability and improving its performance. The effective integration of these methods has also Achieved notable improvement in the field of medical image processing.

While the deep UNet model, which incorporates attention mechanisms, has made substantial strides in medical image processing and analysis, there are still some exploreable issues in multi-attention fusion, inter-layer relationships of UNet, and multi-attention mechanism architecture. Therefore, how to effectively integrate attention mechanisms with deep UNet neural networks, build a multi-layer deep feature fusion computing model, extract the essential features of medical images, and realize the learning of global and regional characteristics, key and auxiliary non-key features, and deep features at different levels has become a hot topic in current research.

2. RELATED WORK

Xing et al. [22] proposed a U-Net framework that integrates multi-attention and multi-scale feature fusion for more effective tumor image segmentation. The algorithm integrates the U-Net, FPN algorithms, and a fused attention mechanism based on channel and pixel-level attention mechanisms, addressing the issue of brain tumor 3D image segmentation and achieving good performance in 3D image segmentation. Li et al. [23]. proposed a novel framework of lung nodule segmentation on CT image that incorporates attention mechanisms and edge detection. The algorithm introduces an information fusion attention module to optimize feature recombination and utilization through attention mechanisms and dilated convolutions. It also integrates the information fusion attention module, edge detection operator, and U-Net model to address the issues of edge definition blur and insufficient segmentation accuracy in CT images, thereby improving segmentation performance. Zhang et al. [24] proposed an attention-based network for enhancing the segmentation performance and accuracy of the model. He proposed the DASGC attention module, which integrates multi-scale spatial features and inter-channel information, and integrated it with U-Net to construct an improved U-Net network (DASGC-U-Net). This approach addresses the issue that U-Net cannot effectively utilize limited image information for precise segmentation, Achieve higher segmentation efficiency in medical image segmentation. Zhong et al. [25] presented a Channel Spatial Attention Nested U-Net for the detection of small targets in infrared images. They developed an innovative architecture, the Channel Spatial Attention Nested U-Net (CSAN-UNet), developed to address the challenges of detecting and segmenting small infrared targets, their research showed outstanding performance. Li et al. [26] proposed Dual Multi-Scale Attention makes U-Net stronger for medical image segmentation. They constructed a self-attention-based Dual Multi-Scale Attention (DMSA) and integrated it with the U-Net network, presenting the DMSA-U-Net model. Experimental results show that this algorithm outperforms other state-of-the-art methods that do not require any pre-trained models. Chen et al. [27] proposed a multi-scale channel attention UNet (MSCA-UNet) to improve the

segmentation accuracy of medical ultrasound images. Experiments show that our method outperforms state-of-the-art (SOTA) methods in accuracy across four medical ultrasound image datasets. Al Qurri and Almekkawy [28] proposed an improved U-Net with attention for medical image segmentation. In the algorithm, a three-level attention framework (TLA) is constructed, which incorporates attention gates, channel attention, and spatial normalization. It integrates the TLA module, TransNorm spatial attention, and the U-Net++ network module, Tackling the issue of the original algorithm's limitation in capturing long-range dependencies. Meng et al. [29] proposed an attention-fused full-scale CNN-Transformer U-Net for medical image segmentation. The attention-fused full-scale CNN-Transformer U-Net (AFC-UNet) aims to effectively overcome the limitations of traditional U-Net through multi-scale feature fusion, attention mechanisms, and a CNN-Transformer hybrid module. Experimental results show that the proposed algorithm achieves better segmentation performance. To address the COVID-19 CT issue, Liu et al. [30] proposed a dedicated segmentation network based on an attention mechanism. The algorithm integrates spatial and channel attention into the U-Net encoder to capture more visual layer information, enabling the identification of normal pixels between adjacent lesion areas. By utilizing a composite function, DTVLoss, which focuses on lesion-area pixels, it addresses issues such as boundary blurring and low contrast caused by the use of BCE in traditional U-shaped networks. Experimental results demonstrate that the proposed algorithm significantly outperforms SOTA COVID-19 segmentation networks. Wang et al. [31] proposed an approach for anterior mediastinal nodular lesion segmentation from chest computed tomography imaging using a U-Net-based neural network with attention mechanisms. The algorithm integrates self-attention mechanisms, convolutional block attention modules (CBAM), and the U-Net network to address the challenge of detecting mediastinal nodular lesions. Experimental results show that the proposed algorithm outperforms related methods. Wu et al. [32] proposed a Multi-scale Efficient Transformer Attention U-Net for fast and high-accuracy polyp segmentation. The algorithm introduces a Multi-scale Efficient Transformer Attention (META) mechanism for adaptive feature fusion, utilizing efficient transformer blocks to generate multi-scale element attention within the renowned U-shaped encoder-decoder architecture. The method demonstrates strong capability in polyp segmentation. Chen et al. [33] proposed a full-scale connected attention-aware U-Net for CT image segmentation of the liver. They developed the Attention UNet3+ algorithm, which outperforms other improved U-Net liver image segmentation algorithms by at least 2.9% in intersection over union (IoU) and at least 1.1% in Dice coefficient.

In summary, the combination of attention mechanisms and deep U-Net networks has gained considerable attention in the field of medicinal image analysis and processing, showing strong progress in theory, technology, and real-world applications. However, the following shortcomings still exist:

- (1) The problem of insufficient segmentation of small targets in UNet: The downsampling operation in UNet may lead to the loss of feature information for small targets (such as micro lesions or fine anatomical structures), affecting the segmentation performance.

- (2) In medical image segmentation with complex backgrounds or regions of similar intensity (such as

neighboring tissues with similar intensity), UNet may suffer from incorrect segmentation.

(3) In medical images, the lesion area is usually small, and the data imbalance can cause the model to be more inclined to predict the background region.

Based on this, this study is proposed. The algorithm mainly achieves the following innovations:

(1) Addressing the Deficiency of U-Net in Small Object Segmentation, this paper improves the coordinate-channel attention mechanism, forming a new channel-coordinate fusion attention mechanism to address attention fusion in three directions: horizontal, vertical, and channel. It constructs the spatial dependencies of the image, enhancing the attention mechanism's focus on the nodules.

(2) U-Net May Encounter Mis-segmentation Issues in Complex Backgrounds or Regions with Similar Intensity. Based on this, a fusion method of self-attention mechanism and channel-coordinate attention mechanism is used to address this issue.

(3) Addressing the Issue of Data Imbalance in Medical Imaging, based on this, the project introduces an improved Tversky loss to enhance the model's sensitivity to the minority class.

3. ALGORITHM DESIGN

3.1 Channel-coordinate attention fusion algorithm

To address the issue of incomplete feature data dependency, this paper proposes a Coordinate-Channel Fusion Attention Mechanism, aiming to construct feature data dependency in a three-dimensional space and enhance the algorithm's focus on the target task. The algorithm begins by calculating the horizontal and vertical spatial projections of the feature data

and applying self-attention mechanisms to these projections to determine the data correlations in the two coordinate-based spaces. Subsequently, a linear extension of the horizontal projection is computed to construct a correlation matrix for feature images in the horizontal and vertical spaces.

Additionally, a channel-wise maximum pooling method is employed to create a channel-wise maximum feature plane (matrix), and a sub-attention relationship matrix is calculated for the maximum feature plane to establish a data correlation matrix in the channel dimension. The channel-wise data correlation matrix is then combined with the feature image correlation matrix in the horizontal and vertical spaces through matrix multiplication, performed coordinate by coordinate across channels. This process calculates the data dependency of the feature matrix in the horizontal, vertical, and channel dimensions, thereby addressing the issue of incomplete data dependency and enhancing the algorithm's focus on the target task.

The specific implementation process is expressed as:

$$Map(out) = (SAM(Project(map(in), x)) \times SAM(Project(map(in), y))) \cdot (SAM(MaxPool(map(in), c))) \cdot Map(in)) \quad (1)$$

Based on the above design concept and inspired by the parallel processing model of visual information in primates, the channel-coordinate attention fusion algorithm adopts two parallel attention-based information processing modes. Therefore, its basic flowchart is as follows.

In Figure 1, the operator \otimes represents vector multiplication, the operator \odot denotes element-wise multiplication between a matrix and a sequence, and the operator \ast signifies element-wise multiplication between matrices within two sequences. Therefore, the pseudo code for this algorithm is as follows.

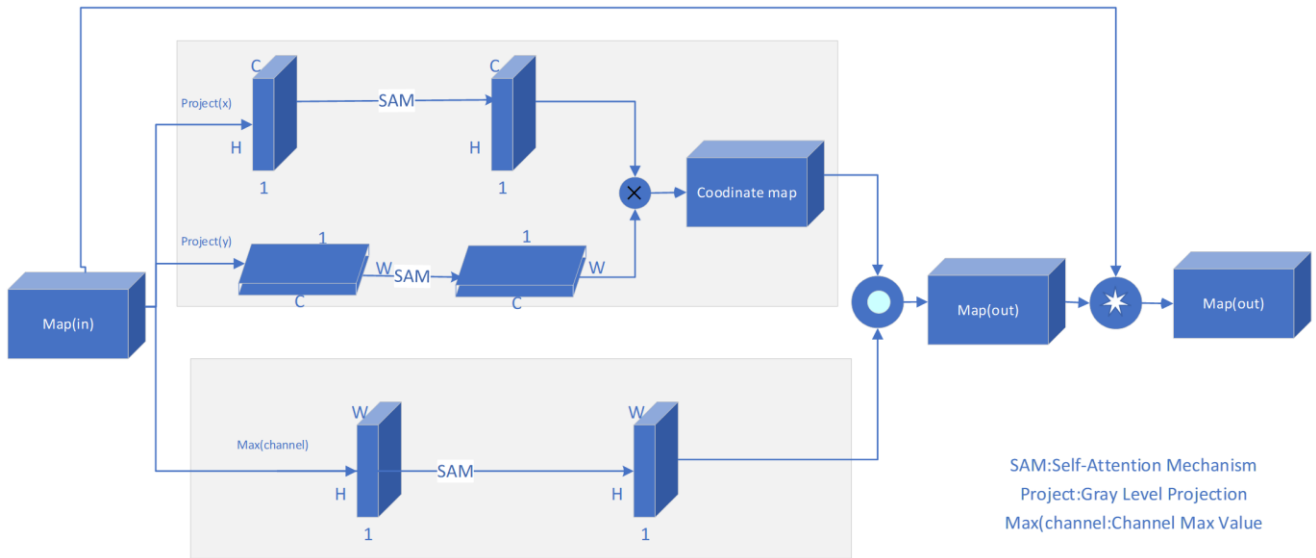


Figure 1. Flowchart of the channel-coordinate attention fusion algorithm

Therefore, the coordinate attention based on the self-attention mechanism can construct the long-range dependencies of feature data elements in the two-dimensional space of the horizontal and vertical spaces; the channel attention mechanism based on the self-attention mechanism constructs the long-range dependencies in the channel dimension of the feature data set; and the fusion of both

constructs the long-range dependencies of feature data elements in the three-dimensional space of the horizontal, vertical, and channel dimensions. As a result, this addresses the issue of incomplete feature dependency and subsequently resolves the problem of small lesion target loss in medical images with complex backgrounds. The pseudo code of channel-coordinate attention fusion algorithm is shown in

Algorithm 1.

Algorithm 1. The pseudo code of channel-coordinate attention fusion algorithm

Input: Map(in)

Produce:

(1) Feature Data Gray Projection.

Using Eqs. (6) and (7), calculate the gray scale projections of the feature data Map(in)Map(in) in the horizontal and vertical directions. The projection is implemented using the gray scale histogram algorithm.

$$\text{outx} = \text{Project}(\text{map}(\text{in}), x) \quad (2)$$

$$\text{outy} = \text{Project}(\text{map}(\text{in}), y) \quad (3)$$

(2) Horizontal and Vertical Dependency Fusion

Calculate the horizontal and vertical dependencies of the projection data using Eqs. (8) and (9), and perform horizontal and vertical dependency fusion.

$$\text{outx_dependency} = \text{SAM}(\text{outx}) \quad (4)$$

$$\text{outy_dependency} = \text{SAM}(\text{outy}) \quad (5)$$

(3) Channel Attention Calculation

Calculate the maximum projection matrix using Equation (10), and compute the self-attention mechanism of the maximum projection matrix to construct the dependency in the channel space.

$$\text{outc_dependency} = \text{SAM}(\text{MaxPool}(\text{map}(\text{in}, c))) \quad (6)$$

(4) Channel-Coordinate Attention Fusion

$$\text{Map}(\text{out}) = ((\text{outx_dependency} \times \text{outy_dependency}) \odot \text{outc_dependency}) \ast \text{Map}(\text{in}) \quad (7)$$

Output: Map(out)

3.2 Pixel-wise Tversky-dice loss function

The Tversky loss function is designed to address the class imbalance issue in medical image segmentation. It is based on the Tversky Index and allows flexible control over the preference of segmentation algorithms for the foreground and background by adjusting the weights of False Positive (FP) and False Negative (FN). Its mathematical description is:

$$L_{Tversky}(\alpha, \beta) = 1 - \frac{|A \cap B|}{|A \cap B| + \alpha|A - B| + \beta|B - A|} \quad (8)$$

It has shown good application results in medical image segmentation. However, once the hyperparameters are determined, the Tversky loss function controlled by the hyperparameters α and β has the problem of statically adjusting the penalties for false positives and false negatives. To address the difficulty of segmenting small targets like pulmonary nodules in medical image processing, and considering the high intensity characteristics of small targets such as pulmonary nodules in medical imaging, a better weight

is assigned to these pixel points. By adjusting the loss function according to the importance of the pixels, the problem of data imbalance is solved.

This paper proposes a new pixel-wise Tversky-Dice loss function (abbreviated as TDLF) by weighted fusion of the pixel-wise Tversky loss function and the pixel-wise Dice loss function, which enhances the evaluation of segmentation accuracy. The mathematical model for this improvement is described as follows:

$$L_{combined} = \gamma \cdot L_{Tversky} - (1 - \gamma) \cdot L_{Dice} \quad (9)$$

$$L_{Tversky} = 1 - \frac{\sum_i w_i \cdot Tversky(p_i, g_i)}{N} \quad (10)$$

$$w_i = \begin{cases} w_{foreground} & \text{if } g_i = 1 \\ w_{background} & \text{if } g_i = 0 \end{cases} \quad (11)$$

$$L_{Dice} = 1 - \frac{2 \cdot \text{intersection} + \epsilon}{\text{Union} + \epsilon} \quad (12)$$

The hyperparameters include: α , β and γ , where $\alpha + \beta = 1$. Therefore, the weighted fusion of both will effectively address the issues of data imbalance and small target loss, while improving the quality of the classification algorithm.

3.3 Adaptive method for setting hyper parameters of the loss function

The DSC measures the overlap between the segmentation result and the ground truth, with a value closer to 1 indicating a more accurate segmentation result. IoU is more stringent than Dice as it penalizes excessively large predicted regions. Therefore, this paper sets the adaptive hyperparameter γ as:

$$\gamma = \frac{mIoU}{Dice} = \frac{1}{2 - mIoU} \quad (13)$$

Here, the hyperparameter γ focuses on the overlap between the segmentation results and the ground truth while better penalizing excessively large predicted regions. Therefore, this parameter enables the adaptive configuration of the Pixel-wise Tversky-Dice Loss Function, ensuring the adaptability of the loss function.

Here, the Tversky index describes the intersection-over-union of task-related feature regions, effectively addressing the problem of small target segmentation under data imbalance conditions. Dice measures the overlap between the segmentation results and the ground truth labels. Therefore, the hyperparameter γ is introduced to represent both the constraint on false positives and false negatives, as well as the similarity between the prediction set and the annotation set.

3.4 Improved UNet neural network based on channel coordinate attention fusion algorithm

The UNet network framework consists of two core modules: an encoder and a decoder. By integrating the aforementioned Channel Coordinate Attention (CCA) innovation with these two core modules, three network structures are proposed: *CCA_Encode_UNet*, *CCA_Decode_UNet*, and *CCA_Encode_Decode_UNet*. These structures embed the CCA module into the encoder, decoder, and both encoder and decoder of UNet, respectively. The embedding location is set at the final layer of the encoder and decoder to enable long-range dependency computation in

3D space.

Each algorithm incorporates the *TDLF* module to address the issue of data imbalance effectively. Subsequent structure ablation experiments demonstrate that *CCA_Encode_UNet*

exhibits superior performance. Therefore, this study adopts this network model, with the specific structure shown in the Figure 2.

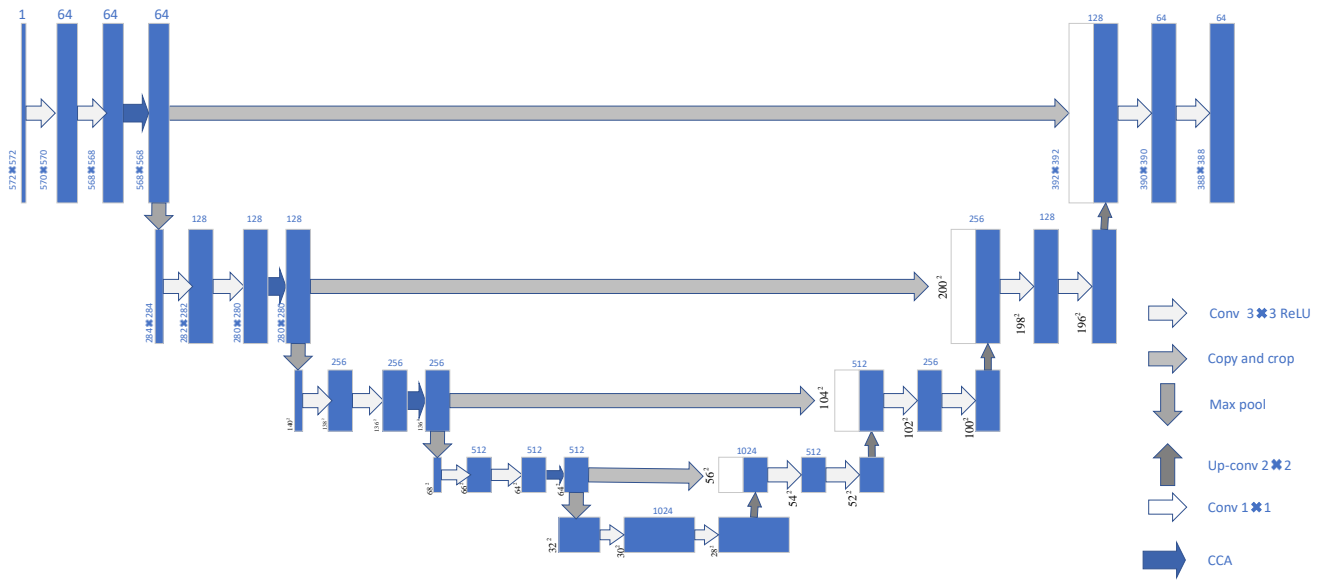


Figure 2. Improved UNet network architecture diagram

4. EXPERIMENTS AND ANALYSIS OF THEIR RESULTS

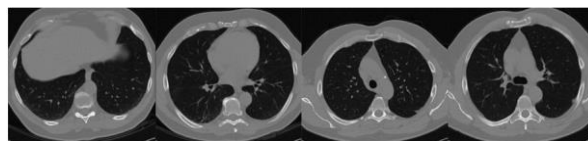
4.1 Introduction to LUNA16

As a subset of LIDC-IDRI, the LUNA16 dataset [34] (Lung Nodule Analysis 2016) is used in this study. It has high academic value and practical application potential. The LUNA16 dataset, which adopts the DICOM data format, does not contain valid lung nodule detection objects in all CT image slices. Therefore, to accelerate the experimental process and ensure the performance of the experiment, preprocessing steps such as image selection and format conversion are necessary.

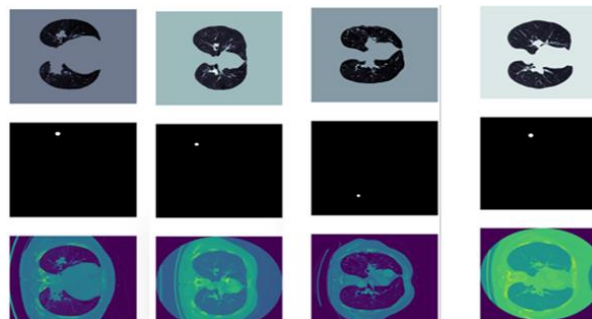
Image selection refers to optimizing and filtering the LUNA16 dataset according to the following criteria. (1) Slice thickness: $0.6 \text{ mm} \leq \text{slice_thickness} < 2.5 \text{ mm}$; (2)

Spatial resolution of slice images: $0.46 \text{ mm} \leq \text{slice_spatial_resolution} \leq 0.98 \text{ mm}$; (3) Nodule size: $\text{Nodule_diameter} \sim 8 \text{ mm}$. Based on these constraints, 888 valid CT images were selected to form the experimental dataset, which contains a total of 1186 positive nodules annotated by four independent radiologists. The images of the dataset contain three dimensions: X, Y, and Z (sequence), which present the complete structure and texture features of the lung nodules from the perspective of sequence space. Image format conversion refers to transforming the DICOM data format into an RGB space format representation. After preprocessing, 888 valid CT images is divided into a test dataset with 950 images and a training dataset with 2850 images, totaling 3800 lung nodule image slices.

Therefore, the LUNA16 dataset samples used in this experiment are shown in Figure 3.



(a) Original dataset



(b) Results of each stage of preprocessing algorithm [35]

Figure 3. The sample of LUNA16

4.2 Innovation of evaluation metrics

To effectively evaluate the performance of the proposed algorithm, five evaluation metrics are introduced: Dice coefficient, IoU, Precision, Recall, and ASD. The specific descriptions are as follows:

Dice Similarity Coefficient

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (14)$$

Here, A represents the segmentation result, and B represents the ground truth label. The Dice Similarity Coefficient (DSC) measures the degree of overlap between the segmentation result and the ground truth. The closer the value is to 1, the more accurate the segmentation result. In medical segmentation, the Dice coefficient is the most commonly used metric.

IoU (Intersection over Union, Jaccard Index)

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (15)$$

Here, IoU represents the intersection-over-union of the predicted region and the ground truth region. Compared to Dice, IoU is more stringent as it penalizes excessively large predicted regions.

Precision

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

Precision reflects the proportion of pixels predicted as the target that actually belong to the target.

Recall, Sensitivity

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

Recall measures the model's ability to detect target regions.

Average Surface Distance, ASD

$$ASD = \frac{1}{|S|} \sum_{s \in S} d(s, T) \quad (18)$$

Here, SS is the predicted surface point set, T is the ground truth surface point set, and $d(s, T)$ represents the distance from point s to the surface T. ASD measures the average distance between the predicted surface and the ground truth surface.

These five metrics construct a more comprehensive evaluation system for medical image segmentation algorithms, evaluating the segmentation performance from various aspects: the overlap between the segmentation result and the ground truth, the intersection-over-union between the predicted and ground truth regions, the proportion of pixels predicted as the target that truly belong to the target, the model's ability to detect the target region, and the average distance between the predicted surface point set and the ground truth surface point set. Therefore, they are used to evaluate the algorithm proposed in this paper.

4.3 Pixel-wise weighted fusion evaluation metrics

To focus on small target regions such as pulmonary nodules, the pixels in the small nodule regions in DICOM imaging typically have higher brightness values. Therefore, brightness

segmentation is performed, and different weight values are assigned to regions based on their brightness. The higher the brightness, the higher the pixel weight for that region. Thus, an adaptive weight calculation method is defined as follows:

$$DSC_{wisepixels} = \frac{2 \times |A \cap B| \times Power_A}{|A| \times Power_A + |B| \times Power_B} \quad (19)$$

$$IoU_{wisepixels} = \frac{|A \cap B|}{|A \cup B|} \times Power_A \quad (20)$$

Therefore, the comprehensive evaluation function is defined as:

$$index = \lambda_1 ASD_{wisepixel} + \lambda_2 IoU_{wisepixels} \quad (21)$$

The DSC measures the overlap between the segmentation result and the ground truth, with a value closer to 1 indicating a more accurate segmentation result. mIoU is more stringent than Dice as it penalizes excessively large predicted regions. Therefore, the comprehensive evaluation metric $index(e)$, which integrates DSC and mIoU can assess both the overlap between the segmentation result and the ground truth, as well as the average distance between the predicted and true surfaces. At the same time, it penalizes excessively large predicted regions. Additionally, it adjusts the segmentation ability of different datasets and algorithms through weighted adjustments.

4.4 Experiments and experimental results analysis

4.4.1 Ablation experiment

Ablation of the attention module: To validate the performance of the channel-coordinate fusion attention, a comparative experiment is conducted based on the UNet network structure using the LUNA16 database. The experiment compares the performance of the UNet algorithm, the coordinate attention-based UNet algorithm, the channel attention-based UNet algorithm, and the channel-coordinate fusion attention-based UNet algorithm. The experimental results of the algorithm comparison are shown in the Figure 4.

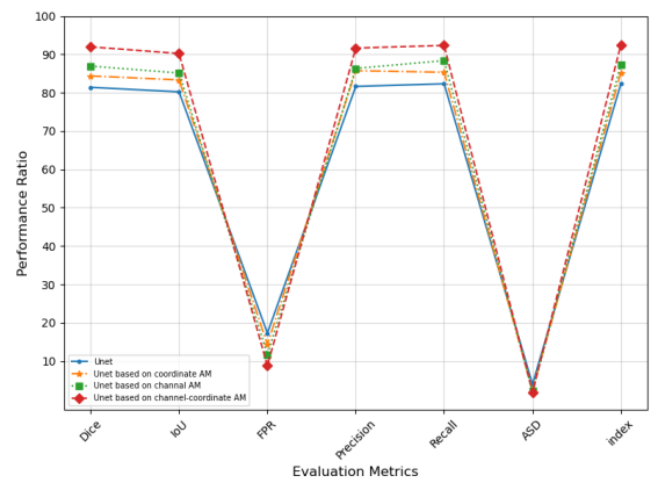


Figure 4. Results of channel and spatial coordinate attention mechanism ablation experiment

Under the same experimental conditions, the experimental results show that the comprehensive segmentation performance of the proposed algorithm is superior to that of

the UNet algorithm that contains only coordinate or channel attention. The comprehensive segmentation performance of the UNet algorithm with either coordinate or channel attention is better than that of the UNet algorithm without any attention mechanism. The reason for this is that both coordinate attention and channel attention can separately learn the long-range dependencies of pixels, enabling focus on small targets like pulmonary nodules in medical images. The fusion of both creates long-range dependencies in 3D space, addressing the issue of insufficient dependencies. Therefore, the segmentation performance of the proposed algorithm is superior to the other three algorithms.

4.4.2 Loss function ablation experiment

To validate the impact of the Tversky-Dice loss function on algorithm performance and confirm the advantages of the proposed improved loss function, a comparative experiment is conducted based on the UNet network structure using the LUNA16 database. The experiment compares the performance of the UNet algorithm with the conventional loss function, the UNet algorithm with the Tversky loss function, the UNet algorithm with the Dice loss function, and the UNet algorithm with the pixel-wise Tversky-Dice loss function. The experimental results of the algorithm comparison are shown in the Figure 5.

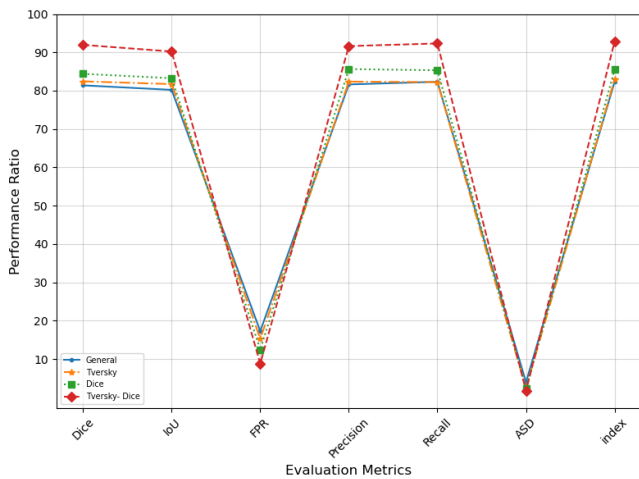


Figure 5. Results of loss function ablation experiment

The experimental results show that the proposed pixel-wise Tversky-Dice algorithm effectively addresses the data imbalance problem in medical image segmentation and enhances the segmentation performance of task-related features. In medical images, small targets such as pulmonary nodules often have higher brightness values, and the algorithm assigns higher weights to these features, highlighting the task-related regions. By adjusting the loss function based on the importance of pixel points, the data imbalance problem is addressed. Additionally, the Tversky-Dice algorithm resolves the static adjustment issue of false positive and false negative penalties in the Tversky loss function. Therefore, the pixel-wise Tversky-Dice algorithm demonstrates better loss computation and data balancing capabilities.

4.4.3 Structural improvement ablation experiment

The CCA module, the pixel-wise Tversky – Dice algorithm, and the two core modules of UNet are fused into

three network structures: CCA_Encode_UNet_TD , CCA_Decode_UNet_TD , and CCA_Encode_Decode_UNet_TD. To validate the performance of the three UNet network structures and confirm the advantages of the proposed algorithm, a comparative experiment is conducted on the LUNA16 database. The experimental results of the three algorithm comparisons are shown in the Figure 6.

The experimental results show that under the same experimental conditions, the image segmentation performance of the three networks CCA_Encode_UNet_TD , CCA_Decode_UNet_TD, and CCA_Encode_Decode_UNet_TD is all good. The segmentation performance of CCA_Encode_UNet_TD is close to that of CCA_Encode_Decode_UNet_TD, and better than that of CCA_Decode_UNet_TD . However, the time and space complexity of CCA_Encode_Decode_UNet_TD is higher than that of CCA_Encode_UNet_TD . Therefore, the CCA_Encode_UNet_TD network is adopted in this paper.

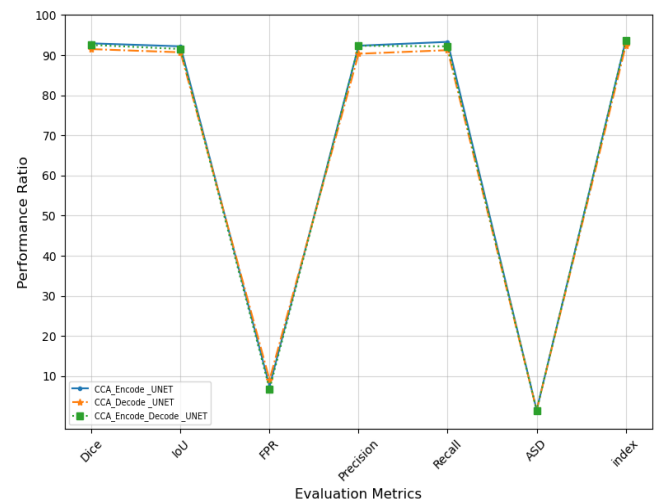


Figure 6. Results of ablation experiment on structural improvement based on CCA method

4.4.4 UNet network layer ablation experiment

The segmentation performance of the UNet network is strongly correlated with the number of layers in the network. As the number of layers increases, the feature learning ability of the UNet algorithm gradually improves, and the segmentation performance of the algorithm increases. However, the time-space complexity of the algorithm rises rapidly, and small target tasks may be lost, leading to a potential decline in algorithm performance. Therefore, the relationship between the number of layers and performance can be determined through an ablation experiment. The results of the ablation experiment are shown in the Figure 7.

On the same experimental conditions, the results of experiments indicate that as the number of layers of the UNet network increases, the algorithm's performance gradually improves and then stabilizes. However, the size of parameters and inference time of the algorithm gradually increase. Therefore, when the performance is similar, the time-space complexity differences of the algorithm become significant. In conclusion, this paper chooses a 5-layer UNet network structure, which ensures both the segmentation performance and the efficiency based on time-space complexity.

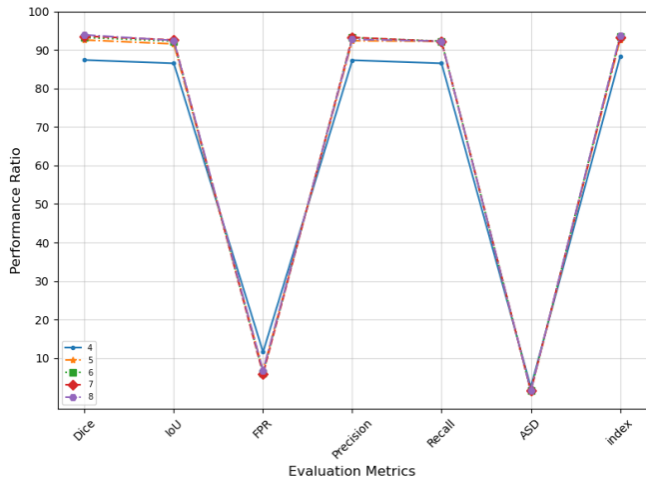


Figure 7. Results of ablation experiment on UNet-like network layers

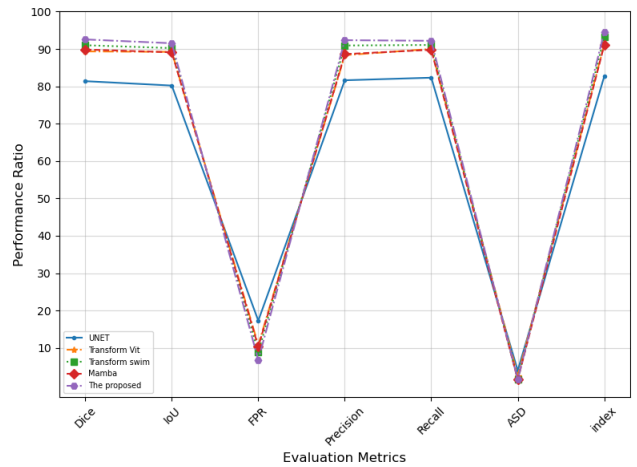


Figure 9. Experimental results of performance comparison across different types of algorithms

4.4.5 Algorithm performance comparison study

To verify the performance advantages of the proposed algorithm, comparative experiments with similar and different segmentation algorithms were conducted on the LUNA16 dataset under the same hardware environment. The experimental results are as follows: In the similar segmentation experiments, the segmentation performance of the proposed algorithm was compared with that of the current UNet, MM-UNet, DASGC-UNet, AFC-UNet, and META-UNet algorithms on the same dataset. The results of the performance experiment are shown in the Figure 8 and Figure 9.

The results of the experiments indicate that, among the similar segmentation algorithms, the algorithm proposed in the paper outperforms the other five common attention-based UNet algorithms, with a significant performance advantage. The reason for this is that the four improvements in the proposed algorithm optimize the UNet algorithm from different aspects, addressing issues such as small target loss and data imbalance in the UNet algorithm.

In experiments comparing different category segmentation algorithms on the same dataset, the proposed algorithm achieved better segmentation performance than the current UNet, Transform Vit, Transform Swin, Mamba, and other algorithms.

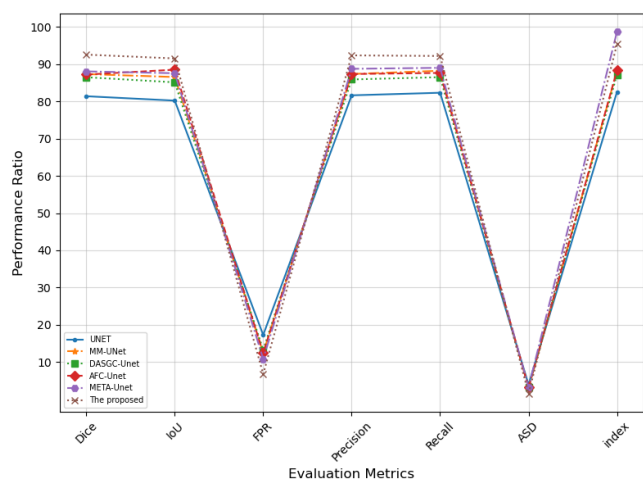


Figure 8. Experimental results of performance comparison of similar algorithms based on UNet network

Experimental results indicate that, in comparison to the current state-of-the-art image segmentation algorithms, the proposed algorithm outperforms others in segmentation performance. Therefore, the proposed algorithm demonstrates a significant competitive advantage in segmentation.

5. CONCLUSION AND OUTLOOK

This paper presents "UNet Based on an Improved Coordinate Channel Attention Mechanism and Its Applications." The proposed algorithm (or model) enhances the coordinate channel attention mechanism by constructing a fused attention mechanism across horizontal, vertical, and channel dimensions. This establishes a three-dimensional dependency in the image, improving the attention mechanism's focus on nodules. Additionally, the sub-attention mechanism improves the coordinate channel attention, addressing misclassification issues in complex backgrounds or regions with similar intensities. The Tversky loss function was enhanced to increase the model's sensitivity to minority classes, resolving issues in UNet's medical image processing, such as small lesion omission, misclassification of highly similar tissues, and data imbalance.

Experiments performed on the LUNA16 dataset show that the algorithm presented in this paper is both feasible and offers better performance compared to conventional algorithms.

In the future, we aim to conduct further in-depth research in the following areas:

- (1) Performance Validation Across Different Datasets: To evaluate the algorithm's generalizability.
- (2) Optimization of Hyperparameter Settings: To further improve algorithm performance.

ACKNOWLEDGMENT

This work is supported by The Introduce intellectual resources Projects of Hebei Province of China in 2023 (The depth computing technology of double-link based on visual selectivity, Grant No.: 2060801); The Introduce intellectual resources Projects of Hebei Province of China in 2025 (3D Point Cloud Segmentation Technology Based on Graph Convolution Algorithm; Key Technologies of Audio

Generation Based on Improved Adversarial Neural Networks); The Key R&D Projects in Hebei Province of China (Grant No.: 19210111D); The Special project of sustainable development agenda innovation demonstration area of the R&D Projects of Applied Technology in Chengde City of Hebei Province of China (Grant Nos.: 202205B031, 202205B089, 202305B101, 202404B104); Higher Education Teaching Reform Research Projects of National Ethnic Affairs Commission of the People's Republic of China in 2021 (Grant No.: 21107, 21106); Wisdom Lead Innovation Space Projects (Grant No.: HZLC2021004).

REFERENCES

- [1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021). 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. <https://doi.org/10.3322/caac.21660>
- [2] Lu, H., Jiang, Z. (2018). Advances in antibody therapeutics targeting small-cell lung cancer. *Advances in Clinical & Experimental Medicine*, 27(9): 1317. <https://doi.org/10.17219/acem/70159>
- [3] Thawani, R., McLane, M., Beig, N., Ghose, S., Prasanna, P., Velcheti, V., Madabhushi, A. (2018). Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer*, 115: 34-41. <https://doi.org/10.1016/j.lungcan.2017.10.015>
- [4] Siegel, R.L., Miller, K.D., Goding Sauer, A., Fedewa, S.A., Butterly, L.F., Anderson, J.C., Jemal, A. (2020). Colorectal cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(3): 145-164. <https://doi.org/10.3322/caac.21601>
- [5] Zheng, K., Wang, X., Jiang, C., Tang, Y., Fang, Z., Hou, J., Hu, S. (2021). Pre-operative prediction of mediastinal node metastasis using radiomics model based on 18F-FDG PET/CT of the primary tumor in non-small cell lung cancer patients. *Frontiers in Medicine*, 8: 673876. <https://doi.org/10.3389/fmed.2021.673876>
- [6] Li, W., Wang, X., Zhang, Y., Li, X., Li, Q., Ye, Z. (2018). Radiomic analysis of pulmonary ground-glass opacity nodules for distinction of preinvasive lesions, invasive pulmonary adenocarcinoma and minimally invasive adenocarcinoma based on quantitative texture analysis of CT. *Chinese Journal of Cancer Research*, 30(4): 415-424. <https://doi.org/10.21147/j.issn.1000-9604.2018.04.04>
- [7] Myers, R., Mayo, J., Atkar-Khattra, S., Yuan, R., Yee, J., English, J., Lam, S. (2021). MA10. 01 prospective evaluation of the International Lung Screening Trial (ILST) protocol for management of first screening LDCT. *Journal of Thoracic Oncology*, 16(10): S913-S914. <https://doi.org/10.1016/j.jtho.2021.08.158>
- [8] Lim, K.P., Marshall, H., Tammemägi, M., Brims, F., McWilliams, A., Stone, E., Lam, S. (2020). Protocol and rationale for the international lung screening trial. *Annals of the American Thoracic Society*, 17(4): 503-512. <https://doi.org/10.1513/AnnalsATS.201902-102OC>
- [9] Shen, D., Wu, G., Suk, H.I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19(1): 221-248. <https://doi.org/10.1146/ANNUREV-BIOENG-071516-044442>
- [10] Lin, T.C., Tsai, C.H., Shiau, C.K., Huang, J.H., Tsai, H.K. (2024). Predicting splicing patterns from the transcription factor binding sites in the promoter with deep learning. *BMC Genomics*, 25: 830. <https://doi.org/10.1186/s12864-024-10667-7>
- [11] Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5): 1122-1131. <https://doi.org/10.1016/j.cell.2018.02.010>
- [12] Kabeya, Y., Okubo, M., Yonezawa, S., Nakano, H., Inoue, M., Ogasawara, M., Nishino, I. (2020). A deep convolutional neural network-based algorithm for muscle biopsy diagnosis outperforms human specialists. *medRxiv*. <https://doi.org/10.1101/2020.12.15.20248231>
- [13] Barrowclough, O.J., Muntingh, G., Nainamalai, V., Stangeby, I. (2021). Binary segmentation of medical images using implicit spline representations and deep learning. *Computer Aided Geometric Design*, 85: 101972. <https://doi.org/10.1016/J.CAGD.2021.101972>
- [14] Le, T.K., Comte, V., Darcourt, J., Razzouk-Cadet, M., Rollet, A.C., Orlhac, F., Humbert, O. (2024). Performance and clinical impact of radiomics and 3D-CNN models for the diagnosis of neurodegenerative parkinsonian syndromes on 18F-FDOPA PET. *Clinical Nuclear Medicine*, 49(10): 924-930. <https://doi.org/10.1097/RLU.0000000000005392>
- [15] Zhou, Y., Zheng, Z., Sun, Q. (2023). Texture pattern-based bi-directional projections for medical image super-resolution. *Mobile Networks and Applications*, 28(5): 1964-1974. <https://doi.org/10.1007/s11036-023-02166-y>
- [16] Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J. (2016). RETAIN: Interpretable predictive model in healthcare using reverse time attention mechanism. *arXiv Preprint arXiv:1608.05745*. <https://doi.org/10.48550/arXiv.1608.05745>
- [17] Qi, Y., Wang, Y., Dong, Y. (2024). Face detection method based on improved YOLO-v4 network and attention mechanism. *Journal of Intelligent Systems*, 33(1): 20230334. <https://doi.org/10.1515/jisys-2023-0334>
- [18] Agrawal, M., Jalal, A.S., Sharma, H. (2024). Enhancing visual question answering with a two-way co-attention mechanism and integrated multimodal features. *Computational Intelligence*, 40(1): e12624. <https://doi.org/10.1111/coin.12624>
- [19] Wei, X., Ye, F., Wan, H., Xu, J., Min, W. (2023). TANet: Triple attention network for medical image segmentation. *Biomedical Signal Processing and Control*, 82: 104608.
- [20] Liao, S., Wang, B., Lin, S. (2024). Optimizing cardiovascular image segmentation through integrated hierarchical features and attention mechanisms. *Technology and Health Care*, 32(1_S): 403-413. <https://doi.org/10.3233/THC-248035>
- [21] Cai, W., Zhai, B., Liu, Y., Liu, R., Ning, X. (2021). Quadratic polynomial guided fuzzy C-means and dual attention mechanism for medical image segmentation. *Displays*, 70: 102106. <https://doi.org/10.1016/j.displa.2021.102106>
- [22] Xing, Y., Yuan, J., Liu, Q., Peng, S., Yan, Y., Yao, J. (2023). MM-UNet: Multi-attention mechanism and multi-scale feature fusion UNet for tumor image segmentation. In *Proceedings of the 2023 2nd Asia Conference on Algorithms, Computing and Machine*

- Learning, pp. 253-257. <https://doi.org/10.1145/3590003.3590047>
- [23] Li, H., Ren, Z., Zhu, G., Wang, J. (2025). AE-UNet: A composite lung CT image segmentation framework using attention mechanism and edge detection. *The Journal of Supercomputing*, 81: 331. <https://doi.org/10.1007/s11227-024-06874-4>
- [24] Zhang, X., Chen, Y., Pu, L., He, Y., Zhou, Y., Sun, H. (2023). DASGC-UNet: An attention network for accurate segmentation of liver CT images. *Neural Processing Letters*, 55(9): 12289-12308. <https://doi.org/10.1007/s11063-023-11421-y>
- [25] Zhong, Y., Shi, Z., Zhang, Y., Zhang, Y., Li, H. (2024). CSAN-UNet: Channel spatial attention nested UNet for infrared small target detection. *Remote Sensing*, 16(11): 1894. <https://doi.org/10.3390/rs16111894>
- [26] Li, X., Fu, C., Wang, Q., Zhang, W., Sham, C.W., Chen, J. (2024). DMSA-UNet: Dual multi-scale attention makes UNet more strong for medical image segmentation. *Knowledge-Based Systems*, 299: 112050. <https://doi.org/10.1016/j.knsys.2024.112050>
- [27] Chen Z, Zhu H, Liu Y, Gao, X. (2024). MSCA-UNet: Multi-scale channel attention-based UNet for segmentation of medical ultrasound images. *Cluster Computing*, 27(5): 6787-6804. <https://doi.org/10.1007/s10586-024-04292-y>
- [28] Al Qurri, A., Almekkawy, M. (2023). Improved UNet with attention for medical image segmentation. *Sensors*, 23(20): 8589. <https://doi.org/10.3390/s23208589>
- [29] Meng, W., Liu, S., Wang, H. (2025). AFC-UNet: Attention-fused full-scale CNN-transformer UNet for medical image segmentation. *Biomedical Signal Processing and Control*, 99: 106839. <https://doi.org/10.1016/j.bspc.2024.106839>
- [30] Liu, X., Liu, Y., Fu, W., Liu, S. (2023). SCTV-UNet: A COVID-19 CT segmentation network based on attention mechanism. *Soft Computing*, 28: 473. <https://doi.org/10.1007/s00500-023-07991-7>
- [31] Wang, Y., Jeong, W.G., Zhang, H., Choi, Y., Jin, G.Y., Ko, S.B. (2024). Anterior mediastinal nodular lesion segmentation from chest computed tomography imaging using UNet based neural network with attention mechanisms. *Multimedia Tools and Applications*, 83(15): 45969-45987. <https://doi.org/10.1007/s11042-023-17210-y>
- [32] Wu H, Zhao Z, Wang Z. (2023). META-UNet: Multi-scale efficient transformer attention UNet for fast and high-accuracy polyp segmentation. *IEEE Transactions on Automation Science and Engineering*, 21(3): 4117-4128. <https://doi.org/10.1109/TASE.2023.3292373>
- [33] Chen C, Shi J, Xu Z, Wang, Z. (2023). Attention UNet3+: A full-scale connected attention-aware UNet for CT image segmentation of liver. *Journal of Electronic Imaging*, 32(6): 063012. <https://doi.org/10.1117/1.JEI.32.6.063012>
- [34] Setio, A.A.A., Traverso, A., De Bel, T., Berens, M.S., Van Den Bogaard, C., Cerello, P., Jacobs, C. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical Image Analysis*, 42: 1-13. <https://doi.org/10.1016/j.media.2017.06.015>
- [35] Zhao, T., Zhao, Y., Ahn, H. (2024). Pyramid model of parallel fusion attention mechanism based on channel-coordinates and its application in medical images. *Traitement du Signal*, 41(5): 2249-2262. <https://doi.org/10.18280/ts.410503>