




Pyramid Model of Parallel Fusion Attention Mechanism Based on Channel-Coordinates and Its Application in Medical Images



Tianshuai Zhao¹, Yanming Zhao², Hyunsik Ahn^{1*}

¹ Tongmyong University, Busan 608830, South Korea

² Office of Academic Research of Hebei Minzu Normal University, Hebei Minzu Normal University, Chengde 067000, China

Corresponding Author Email: hsahn@tu.ac.kr

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410503>

ABSTRACT

Received: 7 April 2024

Revised: 20 August 2024

Accepted: 6 October 2024

Available online: 31 October 2024

Keywords:

deep learning, convolutional network, residual calculation, attention mechanism, pyramid model

The deep residual network model currently used in the detection of small targets such as pulmonary nodules have the problem of the disappearance of small target features caused by the effective fusion of multiple attention mechanisms and the increase in depth. Based on this, "Research on the Pyramid Model of Parallel Fusion Attention Mechanism Based on Channel-Coordinates and Its Application in Medical Images" is proposed. The model proposes an attention mechanism for channel-coordinate parallel fusion, which combines channel attention and temporal attention in parallel, and forms two expressions of an attention mechanism based on the different coordinate embedding timings of channel attention, solving multiple problems. Attention fusion mode and fusion timing issues; Based on this mechanism, combined with the pyramid feature fusion mode, a pyramid model of a parallel fusion attention mechanism based on channel-coordinates is proposed, and the feasibility of the model is theoretically demonstrated. An experiment was organized on RUNA16. The experimental results show that the two attention models proposed for this problem are feasible, have comparative advantages, and the algorithm is stable.

1. INTRODUCTION

Cancer statistics from global research institutions such as the International Agency for Research on Cancer, the World Health Organization, and the Surveillance, Epidemiology, and End Results program indicate that lung cancer is currently the leading cause of cancer-related deaths worldwide. In 2020, there will be approximately 2.2 million new cases of lung cancer and approximately 1.8 million deaths; the mortality rate exceeds 80%, which is the highest among cancer diseases [1]; however, studies [2, 3] show that only 15% of patients are diagnosed at an early stage of the pathological process, which results in higher cancer mortality, with stage I and II lung cancer having a much better prognosis than stage III or IV lung cancer [4]. Therefore, early screening and reasonable treatment of asymptomatic lung cancer are important means to reduce cancer mortality.

Computed tomography (CT) is a key method for effective and non-invasive image detection of pulmonary nodules. Computed tomography images can locate the location of nodules, assess tumor size, perform morphological analysis and prediction, classify tumor attributes, and life cycle analysis and prediction based on the above factors [5-7]. Therefore, they play an important role in the early detection of lung cancer. The National Lung Screening Trial [8-10] showed that screening with low-dose computed tomography (LDCT) will result in a 20% reduction in lung cancer mortality. However, due to the complex background factors of CT images and the small characteristics of pulmonary nodules

themselves, existing human factors and computer models cannot accurately distinguish benign and malignant nodules [11]. Therefore, it is necessary to improve the CT image analysis model and improve the segmentation and classification technology of pulmonary nodules.

In the field of medical image processing, deep learning technology [12-14] that effectively integrates convolutional neural networks has been widely used and achieved good results [15-18]. Radiomics is an emerging technology used to extract high-dimensional quantitative image features for diagnosis and prognosis [19-21]. Texture features of medical images [22-24] play an important role in medical diagnosis and are an important basis for lung cancer diagnosis. Deep convolutional neural networks and radiomics have made good progress in the classification and prediction of pulmonary nodules and in malignant risk assessment. Therefore, it can effectively learn the following characteristics of medical images: (1) Convolutional computation can learn the local detailed features of medical images and the short-range dependencies between pixels; (2) Deep neural networks can learn higher-level abstract features while overlooking low-level features.

However, in medical image analysis, deep convolutional neural networks also have the following problems: (1) As the depth increases, the gradients of deep convolutional neural networks will vanish, which lead to failure of algorithm training. (2) As the depth increases, small goals will be abstracted away. (3) Convolutional neural networks cannot learn global features and long-distance dependencies. (4)

During the learning process, convolutional neural networks cannot focus on and highlight parts of task-related features, that weaken or ignore parts of features that are not relevant to the task. (5) In medical image processing, because there are no relevant national or international norms or standards, the performance of deep convolutional image processing models heavily relies on subjective label data provided by medical experts.

Solving the above problems, Researchers have integrated ResNet, attention modules, pyramid algorithms, and U-shaped network structures to address the aforementioned issues. Such as: A residual calculation module is introduced into the deep convolutional neural network [25-27] to solve the problem of model training failure caused by gradient disappearance, the studies [28-30] explored the integration of deep learning and convolutional neural networks (CNNs) applied to the classification of lung diseases. Using standardized ResNet-50, ResNet-101, and ResNet-152 networks as examples, the research evaluates the algorithm's performance. Experimental results show that the proposed algorithm has a good recognition rate; the studies [31-33] proposed a decision-level fusion technique to improve the performance of CAD systems in classifying pulmonary nodules. This technique is applied to deep residual neural networks, including ResNet-18, ResNet-50, and ResNet-101 models, achieving significant progress in the diagnosis of pulmonary nodules. and an attention mechanism is introduced into the model [34-36], highlighting objects or parts relevant to the task, and ignoring objects or parts irrelevant to the task; using a multi-attention mechanism fusion to initially solve long-distance problems, strengthening model learning capabilities, and improving model performance; using the pyramid algorithm [37, 38] and The U-shaped network algorithm [39-41] implements multi-feature layer-by-layer detection to solve the problem of losing key features of small targets; it uses global computing technology to learn the global characteristics of the data and solves the problem that the convolutional neural network itself can only learn local features; it uses global Medical imaging database solves the problems of inconsistent standards and excessive human dependence. Literature research shows that in the field of image and medical impact processing, the above methods have made good research progress.

Therefore, deep convolutional neural networks that integrate residual calculation, attention mechanism and feature fusion model have made good research progress in medical image process and analysis.

2. RELATED WORK

At this stage, research on the fusion of deep convolutional neural networks, residual calculation, and attention mechanisms has made significant progress in image analysis and processing, providing important support for medical image research.

Ju et al. [42] proposed a feature adaptive algorithm (AFFAM) based on the fusion attention mechanism. This method adaptively integrates the global attention mechanism and the spatial location attention mechanism, which helps to better and more accurate Complex feature maps are learned algorithmically. Apostolidis et al. [43] proposed an algorithmic modeling that integrates the global attention mechanism and the local multi-head attention mechanism. The model accurately learns frame dependencies of different

granularities and scales, and solves the problem of learning interdependencies between features. Deng et al. [44] proposed an attention fusion method that fuses global information and local information at the channel level, and fuses feature maps of different scales to obtain rich semantic information features. Solved the problem of paying enough attention to the object information of small tumors. Abbasi and Saeedi [45] proposed two novel attention modules, global temporal attention and temporal pseudo-Gaussian enhanced self-attention, and integrated them to learn the rich features of time series and improve the time series classification model based on deep learning. performance. Liu et al. [46] proposed an improved U-Net retinal blood vessel segmentation algorithm. The algorithm introduces an improved position attention module and channel attention module at the jump connection to better divide blood vessels and background, and accurately identify blood vessel boundaries. Huang [47] proposed a stroke segmentation network PCMA-UNet based on a fusion attention mechanism. This mechanism integrates pyramid squeeze attention, coordinate attention and multi-scale attention mechanisms to extract different scale spaces in the feature map. information; learn and strengthen the long-range dependencies in the feature space, highlight the key information of the lesion location; highlight the most significant feature maps at different scales to adapt to the adjustment of the current segmented lesion size. Kim and Lee [48] proposed global position self-attention, and realizes the integration of the global position attention mechanism and the channel attention mechanism to achieve global positioning between spatial elements, between consecutive positions, structural geodesic positions, and between channels. The fusion of ordered semantics and positional dependencies strengthens the model's complex feature learning capabilities. Tadepalli et al. [49] studied the combination of ResNet deep learning model and convolutional block attention module (CBAM) to achieve accurate classification of benign pulmonary nodules and malignant nodules, improve early diagnosis accuracy and improve patient prognosis. Cao et al. [50] proposed a multi-scale detection network for pulmonary nodules based on the attention mechanism, designed the ResSCBlock basic module integrating the attention mechanism, and used it for feature extraction. At the same time, the feature pyramid structure is used for feature fusion, and the multi-scale prediction method solves the detection problem of small-sized nodules that are easily lost. Zhang et al. [51] proposed receptive field attention (RFA) to solve the problem of convolution block attention module (CBAM) and coordinate attention (CA) that only focus on spatial features and ignore the problem of large convolution kernel parameter sharing, cannot completely solve the problem of convolution kernel parameter sharing. Receptive field attention can not only focus on the spatial characteristics of the receptive field, but also provide good attention to the weight of large-size convolution kernels. Seung et al. [52] proposed a lung cancer pathology image classification model that combines deep neural networks and customized self-attention ResNet. It solves the vanishing gradient problem and performs well even when layers are densely packed; the custom self-attention module combines channel attention and spatial attention mechanisms to focus on certain features behind the bottleneck structure. In order to accelerate the reconstruction of compressed sensing magnetic resonance images, Chang et al. [53] proposed a dual-domain deep learning algorithm. The algorithm combines deep learning, residual neural network

and attention mechanism to effectively realize magnetic resonance image reconstruction. In the novel the performance of the method is compared with state-of-the-art direct mapping, single-domain and multi-domain methods and achieves good results.

In summary, the effective integration of residual calculation, attention mechanisms, pyramid algorithms, and deep convolutional neural networks has initially addressed the aforementioned five (or possibly four) issues, achieving good results in theory, technology, and application. However, there are still the following shortcomings: (1) When integrating multiple attention mechanisms, which attention mechanisms are selected? (2) When integrating multiple attention mechanisms, how is the serial or parallel processing method for the selected attention mechanisms determined? (3) Fusion Methods of ResNet and Multiple Attention Mechanisms. (4) In deep convolutional neural networks (CNNs), how to solve the problem of small object loss?

Based on this, <Research on the pyramid model of parallel fusion attention mechanism based on channel-coordinates and its application in medical images> is proposed. The algorithm mainly achieves the following innovations:

(1) In a multi-attention mechanism, choose to fuse channel attention and spatial attention mechanisms to achieve the integration between local and global features, as well as between short-range and long-range dependencies.

(2) In the multi-attention mechanism, compare the research on the serial and parallel working modes and working order of channel-spatial attention.

(3) The integration method of residual neural networks and the already defined multi-attention mechanism.

(4) The problem of small object loss is solved by Utilizing a pyramid method to achieve multi-scale feature fusion.

3. MODEL DESIGN

3.1 Model design ideas

In response to the issues raised in this paper, the channel attention mechanism and coordinate attention mechanism were selected as the basic attention mechanisms for multi-attention mechanism fusion. The fusion mode and fusion order of the channel attention mechanism and the coordinate attention mechanism were studied, and the Channel-Coordinate Parallel Fusion Attention Mechanism (Channel-Coordinate PFAM) and its two models were proposed. This addresses the issues of fusion mode and fusion order and enables the effective integration of global channel attention and local coordinate attention, achieving spatial attention for lung nodules from three dimensions. The study also explored the integration of the Channel-Coordinate PFAM with ResNet. At the micro level, the fusion attention mechanism was introduced into the skip connection module of the ResNet network to achieve attention to microscopic image features. At the macro level, the fusion attention mechanism was introduced into the skip connection module of each ResNet layer, achieving layer-by-layer fusion and focusing on different deep features. At the deep model level, the pyramid algorithm was adopted to achieve feature fusion at different depth scales, constructing multi-scale feature associations for lung nodules and solving the problem of small lung nodule targets being lost as the depth of the deep neural network increases.

Therefore, this paper constructs a Channel-Coordinate Parallel Fusion Attention Mechanism and integrates it with Resnet, forming a Resnet module with a fusion attention mechanism. This not only enables the algorithm to focus on important targets but also addresses the gradient vanishing problem. Additionally, the pyramid algorithm is introduced to establish connections between features at different scales and depths, solving the problem of small object loss in deep features of deep neural networks.

3.2 Theoretical analysis

In image analysis and understanding, detecting small targets is challenging due to their size and the strong abstraction capabilities of deep convolutional neural networks. Therefore, various technical approaches can be used to solve or avoid the issue of small target loss. This paper constructs a deep residual neural network based on a fused attention mechanism and pyramid algorithm to address the problem of small target loss.

3.2.1 Channel coordinate attention mechanism selection

In a deep convolutional neural network, the feature map sequence is constructed along the channel (C) dimension, consisting of feature sequences made up of X-Y dimension feature maps.

(1) Channel attention mechanism:

This mechanism enhances the feature representation of important channels by assigning different weights to the different channels of the feature map, while suppressing the features of unimportant channels, thereby improving the model's performance. The algorithm flowchart is as follows:

Step 1. Input feature map: The input feature map $F \in R_{C \times H \times W}$, where C is the number of channels, H is the height, and W is the width.

Step 2. Global Pooling: Perform global average pooling and global max pooling on the input feature map long the spatial dimensions (i.e., $H \times W$ × $W \times H$), generating two importance vectors that describe the significance of the feature channels.

Average pooling descriptor F_{avg}

$$F_{avg} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{c,i,j} \quad (1)$$

Max pooling descriptor F_{max}

$$F_{max} = \text{Max}_{i,j} (F_{c,i,j}) \quad (2)$$

Step 3. Shared Fully Connected Layer (MLP): Pass the two pooling descriptors mentioned above into a shared multi-layer perceptron (MLP) to obtain the channel-level attention weights.

Calculate the attention weights using a two-layer MLP.

$$s = \sigma(\text{MLP}(F_{avg})) + \text{MLP}(F_{max}) \quad (3)$$

Here, σ is the sigmoid activation function, used to constrain the weights between 0 and 1.

Step 4. Channel Reweighting: Adjust the channel weights by multiplying with the input feature map on a per-channel basis, resulting in the weighted feature map.

$$F' = F \cdot s \quad (4)$$

The channel attention mechanism captures which channels contain more valuable features by analyzing the global information of each channel. This allows the network to focus on those channels that are more important for specific tasks while suppressing irrelevant or redundant channel features, thereby enhancing the performance of the network.

(2) Coordinate attention mechanism:

This is an improved attention mechanism that addresses the limitations of traditional spatial and channel attention mechanisms in capturing long-range dependencies and feature localization. It combines the advantages of channel attention and spatial position encoding. By introducing coordinate information, the model can better capture the correlations between different spatial locations, making it particularly suitable for tasks that require precise localization and detail differentiation. The coordinate attention mechanism primarily considers coordinate information when capturing the global dependencies of feature maps, specifically the relative positional information in the spatial dimensions (Height \times Width). Unlike conventional spatial attention mechanisms, coordinate attention not only considers the importance of each spatial location but also leverages global contextual information along the horizontal and vertical coordinates, allowing for more refined focus on the target areas. The algorithm flowchart is as follows:

Step 1. Input Feature Map and Global Pooling: The input feature map size is $H \times W \times C$. Global pooling is performed in both the height and width directions to generate two one-dimensional feature maps.

Height direction

$$X_h = \frac{1}{H} \sum_{i=1}^H X(i, w, c) \quad (5)$$

Width direction

$$X_w = \frac{1}{W} \sum_{j=1}^W X(h, j, c) \quad (6)$$

Step 2. D Convolution and Fusion: Perform convolution operations on the two one-dimensional feature maps separately to generate horizontal and vertical attention maps.

$$F_h = \sigma(\text{Conv1}(X_h)) \quad (7)$$

$$F_w = \sigma(\text{Conv1}(X_w)) \quad (8)$$

Step 3. Coordinate Attention Fusion: Fuse the horizontal and vertical attention maps to obtain an attention map that incorporates global context and positional information.

$$A = F_h \odot F_w \quad (9)$$

Step 4. Feature Weighting: Multiply the generated coordinate attention map with the input feature map on a per-channel basis to obtain the enhanced feature map.

$$X' = A \odot X \quad (10)$$

Advantages: Capturing Long-Range Dependencies: The coordinate attention mechanism captures global information in different directions by decomposing the spatial dimensions, allowing for better modeling of long-distance dependencies.

More Accurate Spatial Localization: By introducing coordinate information, the coordinate attention mechanism

can more precisely locate and distinguish different target areas, making it particularly suitable for tasks such as object detection and image segmentation.

High Computational Efficiency: Compared to traditional spatial attention mechanisms, the coordinate attention mechanism decomposes the spatial dimensions, reducing computational complexity and improving the efficiency of the model.

Therefore, effectively integrating channel attention mechanisms and coordinate attention mechanisms allows for focusing on channels that have higher importance in specific tasks while suppressing irrelevant or redundant channel features. It also enables attention to global dependency information across horizontal and vertical coordinates, allowing for more precise focus on target areas. This approach constructs a fusion of spatial local and global features, as well as short-range and long-range dependencies, facilitating the construction of features.

3.3 Channel-convolution fusion attention mechanism and the application of Resnet

Based on the above research ideas, and inspired by the literature [54], this article designs Channel-Coordinate Fusion Attention Mechanism (Abbreviation: CCFAM) and Residual Module Based on Channel-Coordinate Fusion Attention Mechanism.

3.3.1 Residual module based on channel-coordinate attention parallel mechanism

(1) Mathematical representation of residual module-CCAPM

$$F_X = \text{Poll}_{avg} \left(\sigma \left(\text{Conv1} \left(\frac{1}{H} \sum_{i=1}^H F(i, w, c) \right) \right) \right) \times \left(F \cdot \sigma \left(\begin{array}{l} \text{MLP} \left(\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{c,i,j} \right) \\ + \text{MLP} \left(\text{Max}_{i,j} (F_{c,i,j}) \right) \end{array} \right) \right) \quad (11)$$

$$F_Y = \text{Poll}_{avg} \left(\sigma \left(\text{Conv1} \left(\frac{1}{W} \sum_{j=1}^W F(j, w, c) \right) \right) \right) \times \left(F \cdot \sigma \left(\begin{array}{l} \text{MLP} \left(\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{c,i,j} \right) \\ + \text{MLP} \left(\text{Max}_{i,j} (F_{c,i,j}) \right) \end{array} \right) \right) \quad (12)$$

$$\text{out} = \text{ADD} \left(\sigma \left(\text{pool}_{avg} \left(\text{splite}_2 \left(\text{BatchNorm} \left(\text{conv2d} \left(\text{Concat} (F_X, F_Y) \right) \right) \right) \right) \right) \right) \quad (13)$$

$$F_{res} = F + \text{out} \quad (14)$$

(2) Algorithm flowchart of residual module-CCAPM

Based on the mathematical expression of the residual module-CCAPM, the algorithm flowchart for the Residual Module-CCAPM is Figure 1.

(3) Pseudocode for the residual module-CCAPM algorithm

Based on the mathematical expression of the residual module-CCAPM, the pseudocode the residual module-CCAPM are presented in Table 1.

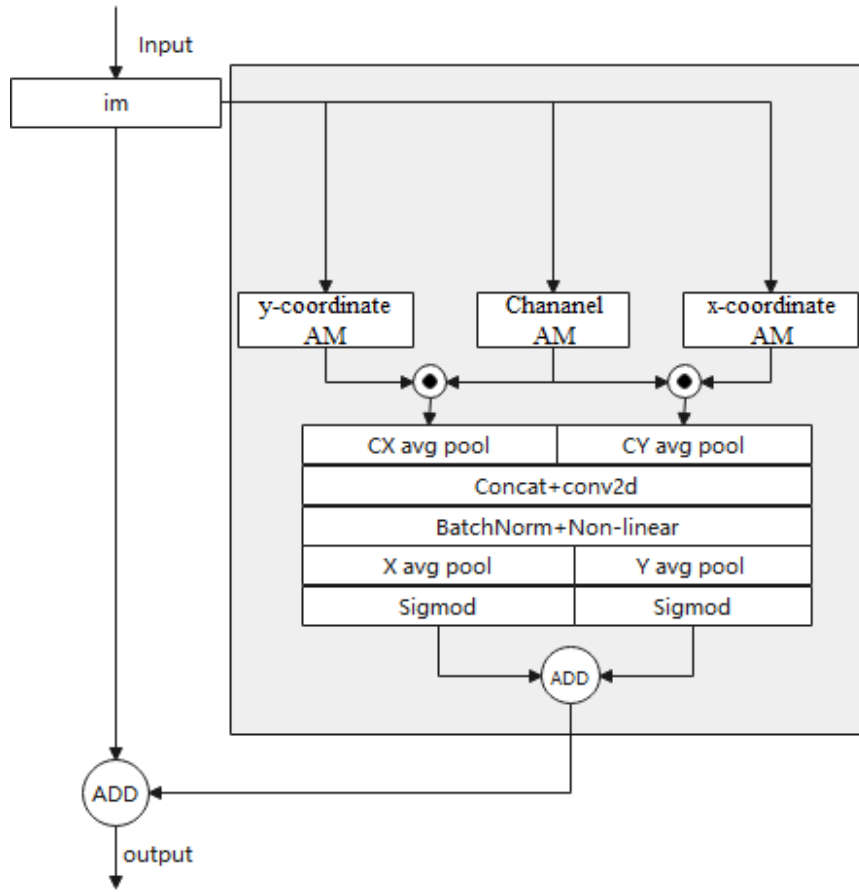


Figure 1. Residual module based on channel-coordinate attention parallel mechanism (residual module-CCAPM)

Table 1. Pseudocode table for the residual module-CCAPM algorithm

<p>Input: The input feature map $F \in R^{(C \times H \times W)}$, where C is the number of channels, H is the height, and W is the width.</p> <p>Procedure:</p> <p>1. According to formula (1) to (8), calculate the channel attention and the X and Y coordinate attention in parallel for the image.</p> $\begin{cases} F_c = F \cdot \sigma \left(MLP \left(\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{c,i,j} \right) + MLP \left(\text{Max}_{i,j} (F_{c,i,j}) \right) \right) \\ F_x = \sigma \left(\text{Conv1} \left(\frac{1}{H} \sum_{i=1}^H F(i, w, c) \right) \right) \\ F_y = \sigma \left(\text{Conv1} \left(\frac{1}{W} \sum_{j=1}^W F(j, w, c) \right) \right) \end{cases}$ <p>2. According to formula (11) and (12), horizontally link the X-coordinate-channel attention and Y-coordinate-channel attention to form the coordinate-channel feature map.</p> $\begin{cases} \bar{F}_y = \text{Poll}_{avg}(F_x) \\ F_y = \text{Poll}_{avg}(F_x) \end{cases}$ <p>3. <i>Concat</i>, <i>conv2d</i>, <i>BatchNorm</i>, <i>splite₂</i>, σ and ADD, etc., on the linked feature map, according to formula (13), sequentially perform operations such as Concat, Conv2D, BatchNorm, Split2, σ, and ADD on the coordinate-channel feature map.</p> $\text{out} = \text{ADD}(\sigma(\text{pool}_{avg}(\text{splite}_2(\text{BatchNorm}(\text{conv2d}(\text{Concat}(F_x, F_y))))))$ <p>4. According to formula (14), calculate the residual result.</p> $F_{res} = F + \text{out}$ <p>Output: F_{Res}</p>
--

3.3.2 Residual module based on channel-coordinate attention serial mechanism

(1) Mathematical representation of residual module-CCASM

$$A = \left(\sigma \left(\text{Conv1} \left(\frac{1}{H} \sum_{i=1}^H F(i, w, c) \right) \right) \right) \times \sigma \left(\text{Conv1} \left(\frac{1}{W} \sum_{j=1}^W F(h, j, c) \right) \right) \quad (15)$$

$$Out = Pool_{avg}((Nonliner(BatchNorm(Conv2d(A)))))) \quad (16)$$

$$Out = F' = \sigma(Pool_{avg}(out \cdot \sigma(MLP(\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W out_{c,i,j}) + MLP(\overset{Max}{i,j}(out_{c,i,j})))))) \quad (17)$$

$$F_{res} = F + out \quad (18)$$

(2) Algorithm flowchart of residual module-CCAPM
Based on the mathematical expression of the residual module-CCASM, the algorithm flowchart for the Residual Module-CCASM is Figure 2.

(3) Pseudocode for the residual module-CCAPSM algorithm
Based on the mathematical expression of the residual module-CCASM, the pseudocode the residual module-CCASM are presented in Table 2.

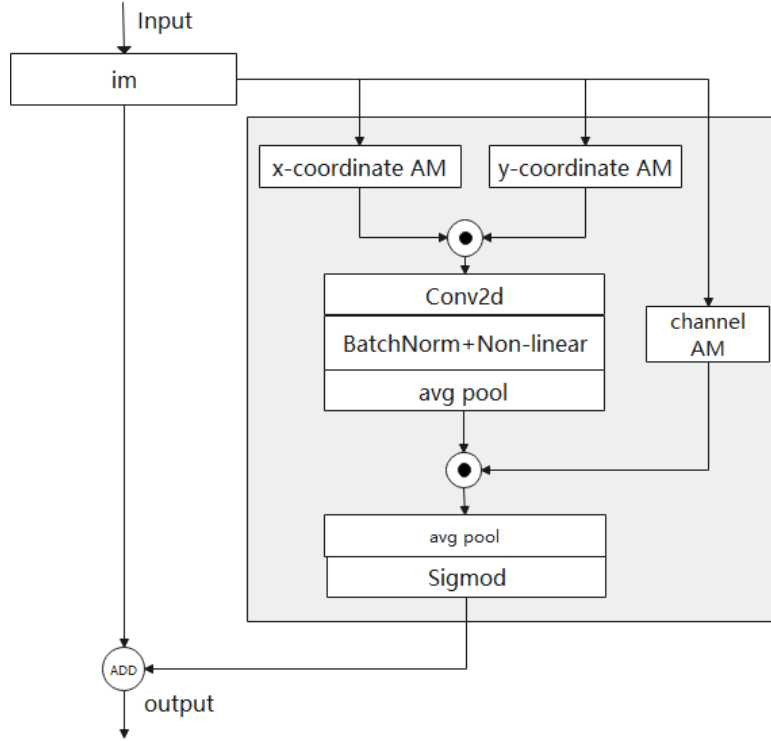


Figure 2. Residual module based on channel-coordinate attention serial mechanism (residual module-CCASM)

Table 2. Pseudocode for the residual module-CCAPSM algorithm

<p>Input: The input feature map $F \in R^{(C \times H \times W)}$, where C is the number of channels, H is the height, and W is the width.</p> <p>Procedure:</p> <ol style="list-style-type: none"> 1. According to formula (15), calculate the X-coordinate attention and Y-coordinate attention of the image in parallel, and perform matrix multiplication. $A = (\sigma(\text{Conv1}(\frac{1}{H} \sum_{i=1}^H F(i, w, c))) \times \sigma(\text{Conv1}(\frac{1}{W} \sum_{j=1}^W F(h, j, c))))$ 2. On the linked feature map, according to formula (16), sequentially perform operations such as conv2d, BatchNorm, Nonlinear, and $Pool_{avg}$ on the coordinate-channel feature map. $out = Pool_{avg}((Nonliner(BatchNorm(Conv2d(A))))))$ 3. According to Eq. (17), the channel attention mechanism is applied to the feature map, resulting in the calculation of: $out = F' = \sigma(Pool_{avg}(out \cdot \sigma(MLP(\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W out_{c,i,j}) + MLP(\overset{Max}{i,j}(out_{c,i,j}))))))$ 4. According to formula (18), calculate the residual result. $F_{res} = F + out$ <p>Output: F_{Res}</p>
--

3.4 Deep residual neural network that integrates channel-coordinate attention mechanism and pyramid algorithm

The Feature Pyramid Network (FPN) algorithm simultaneously leverages the high-resolution detail information from lower layers and the high-level abstract semantic information from upper layers. By fusing these different layer features, it achieves multi-scale feature fusion, addressing the problem of losing small objects. Additionally, predictions are made independently on each fused feature layer, which differs from conventional feature fusion methods.

The Region Proposal Network (RPN) is an efficient and learnable region proposal generation algorithm that greatly accelerates object detection tasks. By integrating with convolutional neural networks, it enables end-to-end training. In lung nodule image segmentation, the role of RPN is to generate Regions of Interest (ROI), where more fine-grained segmentation is performed.

By integrating FPN and RPN, this method can effectively improve the accuracy and efficiency of lung nodule image segmentation. FPN provides the ability to extract multi-scale features, while RPN helps generate candidate regions, allowing the segmentation task to focus on the lung nodule regions of interest. Combined with architectures like Mask R-CNN, this approach can produce high-quality lung nodule segmentation results. This method is particularly effective for small object segmentation in medical images.

Main step of algorithm

Step 1. Input: $M \times N \times N$ image; this paper uses a preprocessed subset of the LUNA16 image dataset, where the image sequence size is $3 \times 512 \times 512$. Through the image sequence preprocessing method, the image sequence is processed into an image sequence of size $32 \times 512 \times 512$, denoted as $I_{32 \times 512 \times 512}$.

Step 2. Feature learning: The Residual-CCFAM (Such as CCASM or CCAPM) module is used to replace the convolutional processing module group of the deep convolutional neural network, constructing a deep convolutional neural network with a depth of 5 to complete deep feature extraction from the input images. Additionally, the feature map sequence at each depth level is preserved.

$$C_i = Residual - CCFAM(C_{i-1}), i = 1, 2, 3, 4, 5 \quad (19)$$

Step 3. Multi-scale feature fusion based on pyramid algorithms: Using pyramid algorithms to fuse features at different depth scales, forming a feature set.

$$P_l = conv_{1 \times 1}(C_i) + upsample(P_{l+1}) \quad (20)$$

Here, P_l is the feature map of each layer in the pyramid; represents a 1×1 convolution applied to the backbone feature map C_i to reduce the channel dimension and enhance the features. $upsample(P_{l+1})$ upsamples the higher-level feature map P_{l+1} to the same resolution as C_i and performs element-wise addition with $conv_{1 \times 1}(C_i)$. This process constructs a series of feature maps from the higher layer P_5 down to the lower layer P_2 , following a top-down approach layer by layer.

Step 4. Semantic segmentation feature generation: Complete two fully connected mappings, and then perform RPN transformation on the mapping results to generate a set of bounding box regression and classification outputs.

The RPN slides a fixed-size window (in this paper, a 3×3

convolutional kernel) over each generated feature map (P_l) to extract local features and generate candidate boxes.

$$RPN(P_l) = \{RPN_{cls}(P_l), RPN_{reg}(P_l)\} \\ = Conv_{3 \times 3}(P_l) \quad (21)$$

For each sliding window position, the RPN generates k anchors, which correspond to different scales and aspect ratios. Assuming each position generates k anchors, the output of the RPN includes:

(1). Classification scores (cls): This represents the probability of each anchor being foreground (object) or background.

(2). Bounding box regression (reg): This indicates the regression offset values for each anchor, used to adjust the coordinates of the anchors to better fit the objects.

The overall structure of the deep residual neural network that integrates the channel-coordinate attention mechanism and pyramid algorithm is shown in Figure 3.

Applying the RPN to feature maps P_l of different scales generates multi-scale candidate boxes. By generating anchors on each feature map P_l , the RPN can produce candidate regions at multiple scales. Thus, the RPN runs simultaneously across multiple scales in the FPN, and its output can be represented as: $\{RPN_{cls}(P_l), RPN_{reg}(P_l)\}$. The classification and regression outputs from each layer correspond to a set of candidate boxes at different scales, which are ultimately refined by applying Non-Maximum Suppression (NMS) to eliminate redundant boxes.

Step 5. NMS Post-processing

Finally, the redundant candidate boxes generated at multiple scales are eliminated through Non-Maximum Suppression (NMS) processing to obtain the final region proposals.

$$Proposals = NMS(\{RPN_{cls}(P_l), RPN_{reg}(P_l)\}, l) \\ = 1, 2, 3, 4, 5 \quad (22)$$

Output: The set of box regressions and classifications is called proposals.

Based on this, the structure diagram of the Deep Residual Neural Network that Integrates Channel-Coordinate Attention Mechanism and Pyramid Algorithm is as follows:

The model of this article is divided into five parts, the input module, the four-layer deep feature learning module based on the attention mechanism of this article, the feature fusion module of the multi-scale pyramid structure, the classification and prediction module and the output module. The input module can use appropriate processing algorithms [55] to generate input sequences according to different databases; the four-layer deep feature learning module based on the attention mechanism of this article implements feature extraction according to depth increments and generates four-scale fusion features; multi-scale In the feature fusion module, after the four scale features are subjected to FP operation, the upsampling method (Upsample) is used to realize the feature map size and then matrix operation is performed to realize four depth feature extraction and solve the layer-by-layer feature learning and fusion of small targets; According to the fused features of four scales, the classification and prediction module implements the classification and prediction of feature maps; the output layer is used to output the splitting results and display them on the original image.

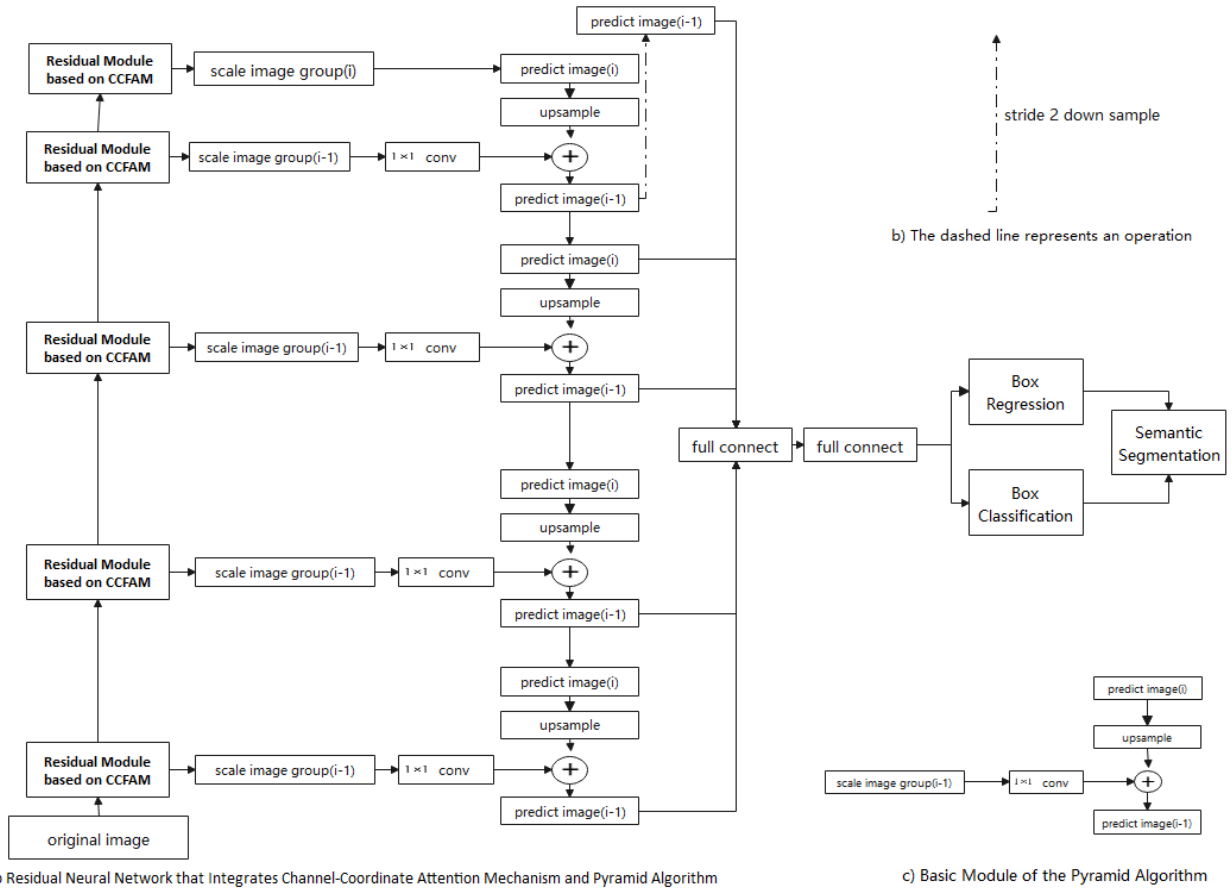


Figure 3. Overall structure of deep residual neural network that integrates channel-coordinate attention mechanism and pyramid algorithm

4. EVALUATION INDEX

The algorithm in this article is designed to solve the detection of small targets such as pulmonary nodules. Therefore, the Mean Average Precision (mAP) method is selected as the evaluation target. mAP comprehensively considers precision and recall, and calculates their average to achieve multiple threshold evaluations, which can more comprehensively evaluate the model's performance in object detection tasks. Often used to handle multi-category target detection.

The recall rate represents the proportion of all positive samples that are correctly predicted. The calculation method is as follows:

$$Recall = \frac{TP}{TP+FN} \quad (23)$$

Precision represents the proportion of all correctly predicted positive samples to all predicted positive samples. Calculated as follows:

$$precision = \frac{TP}{TP+FP} \quad (24)$$

The area value of the PR curve drawn according to the precision rate, recall rate and coordinate axis is the AP value. Since only one category is set, the AP value is equal to the mAP value.

$$\begin{cases} AP_i = \sum_n (Recall_n - Recall_{n-1}) \times precision_n \\ mAP = \frac{1}{C} \sum_{i=1}^C AP_i \end{cases} \quad (25)$$

Jaccard Coefficient (also known as Intersection over Union, IoU) is a metric used to measure the similarity between two sets. It is defined as the size of the intersection of two sets divided by the size of their union:

$$Jaccard\ Coefficient = \frac{A \cap B}{A \cup B} \quad (26)$$

F1 Score is a metric used to evaluate the performance of a classification model, particularly suitable for imbalanced datasets. It is the harmonic mean of Precision and Recall, and it provides a balanced measure of a classifier's accuracy and coverage in predictions.

The formula for calculating the F1 score is:

$$F1\ Score = 2 \times \frac{Recall \times precision}{Recall + precision} \quad (27)$$

5. EXPERIMENTS AND ANALYSIS OF THEIR RESULTS

5.1 Databases

This article uses the LUNA16 [56] (Lung Nodule Analysis 2016) data set derived from LIDC-IDRI (lung image database consortium and image database resource initiative), which is a

subset of the largest public lung CT image database. And press the "Filter out CT image conditions with slice thickness greater than 2.5 mm, select 888 CT images with slice thickness ranging from 0.6 to 2.5 mm, spatial resolution ranging from 0.46 to 0.98 mm, and average diameter of 8.3 mm. Image construction data set. The criterion for pulmonary nodules in the LUNA16 data set is that at least 3 of 4 radiologists believe that the diameter of the nodule is greater than 3 mm, and a total of 1186 positive nodules are labeled. CT sequence, contains three dimensions: each image is the X, Y dimension, and the sequence is the Z dimension. A single CT sequence may contain multiple lung nodules at different locations. The data set is first segmented and then the data is enhanced. as the picture shows.

Since LUNA16 uses the DICOM data format and not every CT slice contains lung nodule content, a custom image processing algorithm is implemented to initialize the LUNA database and generate the experimental dataset. The dataset consists of 3,800 images of lung nodules, which are split into training and testing sets in a 3:1 ratio. Specifically, the training set contains 2,850 images, while the testing set includes 950 images.

Therefore, the LUNA16 dataset and the processed data used in this experiment are shown in Figure 4.

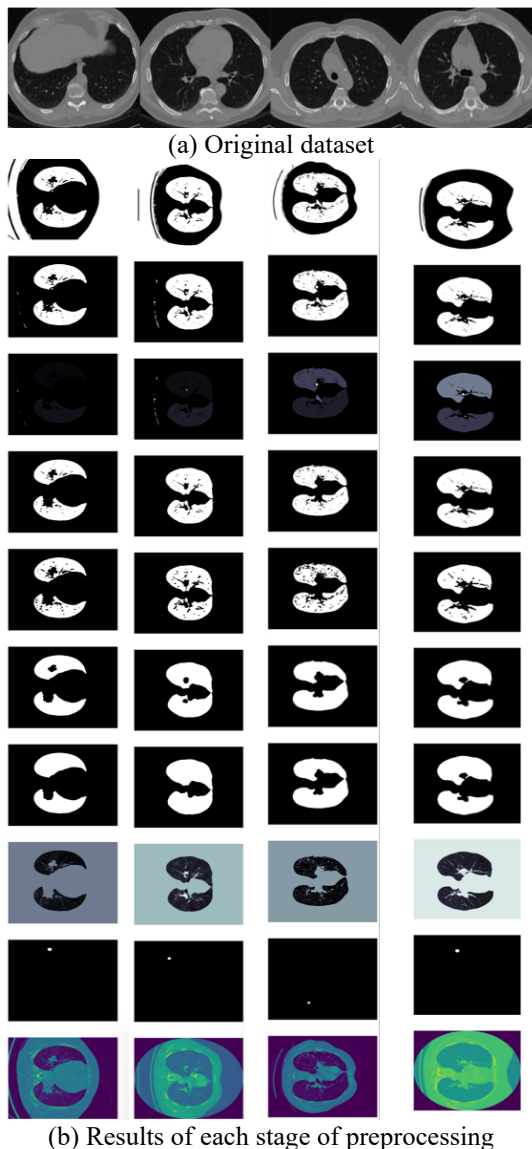


Figure 4. The sample of LUNA16

5.2 Experimental conditions

The main software environment of this experiment includes the Windows 10 operating system, the application development environment Python 3.10 development environment, the PyTorch deep learning framework and its NVIDIA GPU support system; the main hardware includes a dual-channel Intel(R) Xeon Gold 6133 CPU; the memory is 128 GB; 4 NVIDIA GeForce RTX 3090 GPUs with 24 GB of memory. The experiments in this article were conducted on GPU.

5.3 Experiment of this article

This article organizes three groups of experiments to verify the correctness and performance of the parallel fusion attention mechanism of channel and coordinate attention. The first set of experiments uses the Resnet structure as the carrier of the deep neural network to compare the performance advantages of different attention mechanisms and their integration. The second set of experiments compares the pyramid model based on the parallel fusion attention mechanism of channel and coordinate attention with other deep traditional excellent models. Specific experimental process:

5.3.1 Performance verification of the attention mechanism in this article

Using LUNA16 as the data set and the Resnet structure as the carrier of the deep neural network, we compare and verify the performance advantages of the attention mechanism in this article and the different attention mechanisms related to this article. The comparative experimental results are shown in Figure 5 and Table 3.

Using Resnet as the deep learning framework, the experimental results of comparing the attention mechanism of this article with different attention mechanisms related to this article show that the optimal values of mAP and Recall of the two attention mechanisms proposed in this article are better than those listed in this article. force mechanism; as the epoch value increases, the mAP and Recall of the two attention mechanisms proposed in this article first increase, and then gradually become stable, and the overall performance is better than the related algorithms listed in this article. Therefore, the attention mechanism proposed in this article is correct and feasible on deep neural networks and has good performance advantages. The reason is that the SE module fuses local information into global information through compression operations and assigns high weight to target information by reducing the weight of unimportant information in the channel dimension. Therefore, using the SE module alone can improve the detection ability of nodules, but the classification ability of easily confused structures is still insufficient. According to the experimental results, the performance of the ECA (Efficient Channel Attention, ECA) module is slightly inferior to the SE module because the ECA module reduces the complex dependencies between channels through the cross-channel strategy, although a certain degree of channel relationships is retained. However, this strategy did not achieve good detection results in the overall detection process. The CBAM module separates channel attention and spatial attention, but the overall detection effect is poor compared with the SE module and ECA module. The algorithm in this article uses the channel and coordinate attention parallel refuse algorithm

to achieve complete integration of global features and local features, long-distance dependencies and local dependencies.

Therefore, the attention mechanism in this article can achieve better performance.

Comparison table

Table 3. Comparison of detection reflects of different attention mechanisms based on resnet

Method	mAP	Recall	Jaccard Coefficient	F1 Score
Res [40]	0.7765	0.4542	78.16	79.72
Res-SE [41]	0.7785	0.4655	78.59	80.23
Res-Coord [42]	0.7239	0.4284	73.87	75.33
Res-ECA [43]	0.7452	0.4281	75.27	77.37
Res-CBAM [44]	0.6524	0.4187	67.07	68.36
ResSCBlock [45]	0.8328	0.4575	84.11	82.02
Residual Module-CCASM	0.8351	0.4572	84.67	85.31
Residual Module-CCAPM	0.8421	0.4599	85.56	86.43

Comparison chart

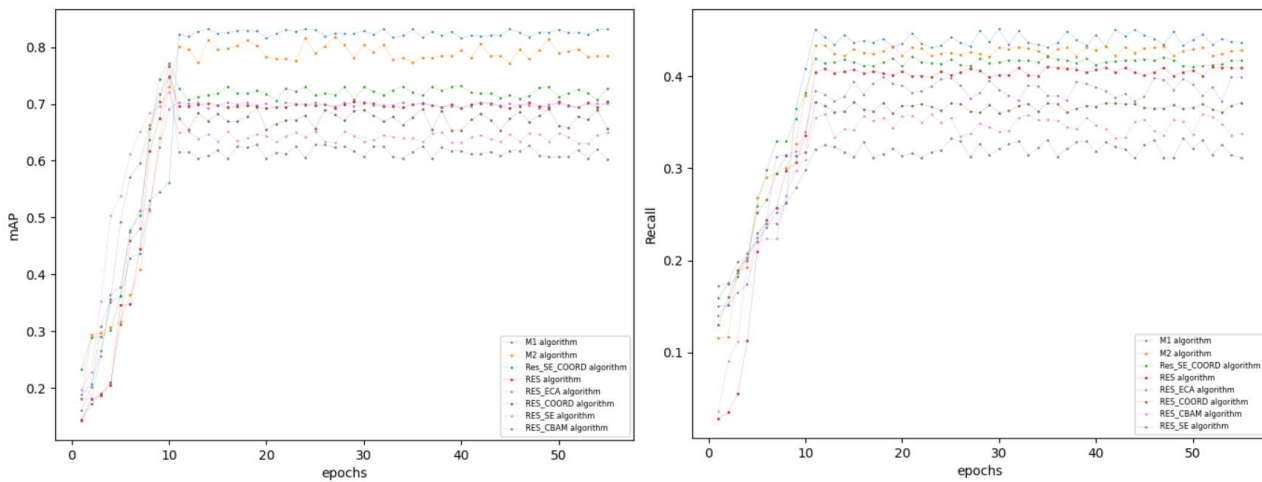


Figure 5. Comparison of mAP and Recall performance of different attention mechanisms

5.3.2 Comparative study of this model and other traditional models

On LUNA16 as the data set, VGG (Fast RCNN architecture), Resnet18, Resnet50, and ResSCBlock were used as comparative research objects to verify the performance comparative advantages of the pyramid model based on the parallel fusion attention mechanism of channel and coordinate attention proposed in this article. The comparative experimental results are shown in Table 4 and Figure 6.

Using Resnet as the deep learning framework, the model proposed in this article compared with VGG (Fast RCNN architecture), Resnet18, Resnet50, and ResSCBlock. The experimental results show that: the mAP and mAP of the pyramid model based on the parallel fusion attention mechanism of channel and coordinate attention proposed in this article are the optimal value of Recall is better than the attention mechanism models listed in this article; as the epoch

value increases, the mAP and Recall of the model proposed in this article first increase, and then gradually become stable, and the overall performance is better than those listed in this article. Related models. Therefore, the model proposed in this article is correct and feasible, and has good performance advantages. The reason is: the algorithm in this article uses the channel and coordinate attention parallel Rongfuhe algorithm to achieve the complete integration of global features and local features, long-distance dependencies and local dependencies. Therefore, the attention mechanism in this article can achieve better performance; this article The pyramid feature fusion mode is used to solve the problem of the disappearance of small target features of pulmonary nodules as the depth increases. Therefore, the pyramid model proposed in this article based on the parallel fusion attention mechanism of channel and coordinate attention has achieved better comparative advantages.

Table 4. Comparison of detection reflects of different model

Method	mAP	Recall	Jaccard Coefficient	F1 Score
Vgg16 [46]	0.6164	0.3820	67.03	67.92
ResNet18 [47]	0.7615	0.4403	77.23	77.89
ResNet50 [48]	0.7915	0.4522	79.13	80.12
Efcientnet-B0 [49]	0.7322	0.4211	74.18	75.61
mobilenet_v3_large [50]	0.6457	0.3879	64.78	65.02
ResSCBlock [51]	0.8308	0.4565	83.23	83.87
Residual Module-CCASM	0.8238	0.4542	82.39	83.15
Residual Module-CCAPM	0.8468	0.4632	84.67	85.78

Comparison chart:

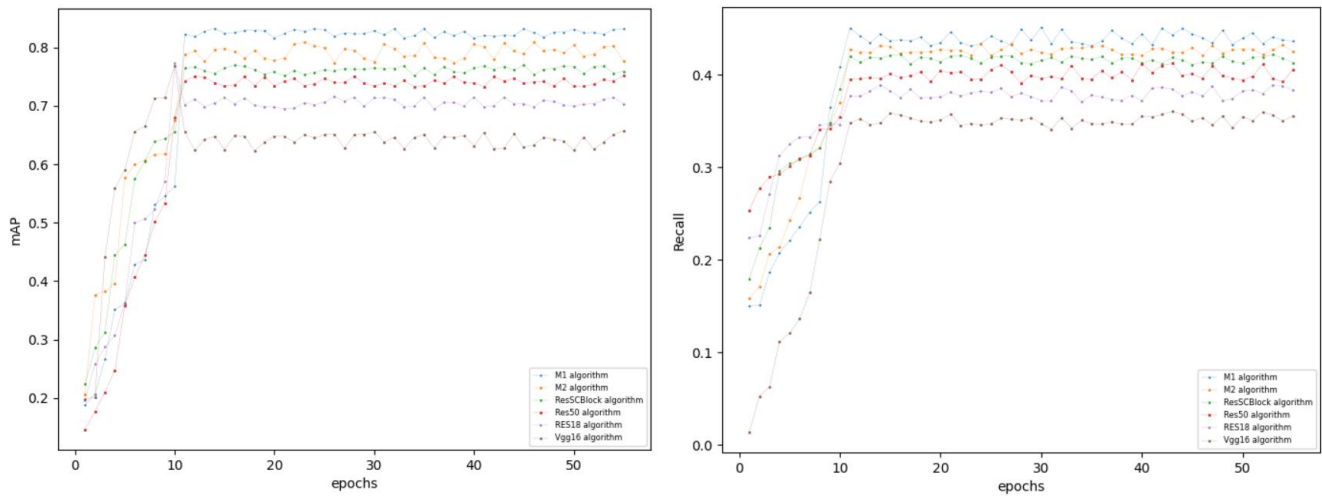


Figure 6. Comparison of mAP and Recall performance of different algorithms

6. CONCLUSIONS

Problems in the current stage of CT image processing for pulmonary nodules. <Research on the pyramid model of parallel fusion attention mechanism based on channel-coordinates and its application in medical images> is proposed. This paper achieves the following innovations: (1) In a multi-attention mechanism, choose to fuse channel attention and spatial attention mechanisms to achieve the integration between local and global features, as well as between short-range and long-range dependencies. (2) In the multi-attention mechanism, compare the research on the serial and parallel working modes and working order of channel-spatial attention. (3) The integration method of residual neural networks and the already defined multi-attention mechanism. (4) The problem of small object loss is solved by Utilizing a pyramid method to achieve multi-scale feature fusion; The Residual Module-CCASM and Residual Module-CCAPM models, which integrate attention mechanisms, were proposed. Based on these integrated attention mechanisms, two deep graph convolution algorithms were constructed to address the aforementioned problems for pulmonary nodules images. The models were validated on the LUNA16 dataset, and the experimental results demonstrated that the proposed algorithms are feasibility. Compared to traditional algorithms, the proposed methods exhibited better comparative performance.

Although the algorithm in this article has made good research progress, it still needs the following improvements:

- (1) Data set size and versatility expansion. Verify this algorithm on a larger and more general medical data set; verify this algorithm on a non-medical image data set.
- (2) Based on the model of this article, design and implement more attention fusion mechanisms to replace the attention mechanism of this article, and verify the feasibility and comparative advantages of the attention mechanism.
- (3) Investigate the parameter settings of the proposed algorithm on the LUNA16 dataset to provide a reasonable parameter setting scheme for algorithm transfer.

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 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); 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] Yu, W., Tang, C., Hobbs, B.P., Li, X., Koay, E.J., Wistuba, I.I., Chang, J.Y. (2018). Development and validation of a predictive radiomics model for clinical

- outcomes in stage I non-small cell lung cancer. *International Journal of Radiation Oncology* Biology* Physics*, 102(4): 1090-1097. <https://doi.org/10.1016/j.ijrobp.2017.10.046>
- [6] Wang, F., Yang, H., Chen, W., Ruan, L., Jiang, T., Cheng, L., Fang, M. (2024). A combined model using pre-treatment CT radiomics and clinicopathological features of non-small cell lung cancer to predict major pathological responses after neoadjuvant chemoimmunotherapy. *Current Problems in Cancer*, 50: 101098. <https://doi.org/10.1016/j.currproblcancer.2024.101098>
- [7] 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>
- [8] 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>
- [9] Tammemägi, M.C., Ruparel, M., Tremblay, A., et al. (2022). USPSTF2013 versus PLCOm2012 lung cancer screening eligibility criteria (International Lung Screening Trial): Interim analysis of a prospective cohort study. *The Lancet Oncology*, 23(1): 138-148. [https://doi.org/10.1016/S1470-2045\(21\)00590-8](https://doi.org/10.1016/S1470-2045(21)00590-8)
- [10] 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>
- [11] 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>
- [12] Ashtagi, R., Khanapurkar, N., Patil, A.R., Sarmalkar, V., Chaugule, B., Naveen, H.M. (2024). Enhancing pneumonia diagnosis with transfer learning: A deep learning approach. *Information Dynamics and Applications*, 3(2): 104-124. <https://doi.org/10.56578/ida030203>
- [13] Rehman, A., Butt, M.A., Zaman, M. (2022). Liver Lesion Segmentation Using Deep Learning Models. *Acadlore Transactions on AI and Machine Learning*, 1(1): 61-67. <https://doi.org/10.56578/ataiml010108>
- [14] Hou, X.X., Liu, R.B., Zhang, Y.Z., Han, X.R., He, J.C., Ma, H. (2024). NC2C-TransCycleGAN: Non-contrast to contrast-enhanced CT image synthesis using transformer CycleGAN. *Healthcraft Frontiers*, 2(1): 34-45. <https://doi.org/10.56578/hf020104>
- [15] 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>
- [16] Sharma, V., Gupta, S.K., Shukla, K.K. (2024). Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images. *Intelligent Medicine*, 4(2): 104-113. <https://doi.org/10.1016/j.imed.2023.06.001>
- [17] 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*, 2020-12. <https://doi.org/10.1101/2020.12.15.20248231>
- [18] 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>
- [19] Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R.G., Granton, P., Aerts, H.J. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4): 441-446. <https://doi.org/10.1016/j.ejca.2011.11.036>
- [20] Zhou, Y., He, L., Huang, Y., Chen, S., Wu, P., Ye, W., Liang, C. (2017). CT-based radiomics signature: A potential biomarker for preoperative prediction of early recurrence in hepatocellular carcinoma. *Abdominal Radiology*, 42: 1695-1704. <https://doi.org/10.1007/s00261-017-1072-0>
- [21] Le, T.K., Comte, V., Darcourt, J., Razzouk-Cadet, M., Rollet, A.C., Orhac, 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>
- [22] Zhou, Z., Chen, K., Hu, D., Shu, H., Coatrieux, G., Coatrieux, J.L., Chen, Y. (2024). Global texture sensitive convolutional transformer for medical image steganalysis. *Multimedia Systems*, 30(3): 155. <https://doi.org/10.1007/s00530-024-01344-6>
- [23] 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>
- [24] Nanditha, B.R., Dinesh, M.S., Murali, S., Chandrasheka, H.S. (2020). Texture analysis of color oral images for lesion detection. In *2020 International Conference on Computational Performance Evaluation (ComPE)*, Shillong, India, pp. 067-072. <https://doi.org/10.1109/ComPE49325.2020.9200023>
- [25] He, K., Zhang, X., Ren, S., Sun, J. (2016). Identity mappings in deep residual networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, the Netherlands*, pp. 630-645. https://doi.org/10.1007/978-3-319-46493-0_38
- [26] Jain, A., Moparthy, N.R., Swathi, A., Sharma, Y.K., Mittal, N., Alhussen, A., Haq, M. (2024). Deep learning-based mask identification system using ResNet transfer learning architecture. *Computer Systems Science & Engineering*, 48(2): 341-362. <https://doi.org/10.32604/csse.2023.036973>
- [27] Echalar, L., Fajardo, A.C. (2023). CNN's Resnet, yolo, and faster R-CNN architectures on the disease and pest classification of local agricultural vegetables towards sustainable production. *SSRN*. <https://doi.org/10.2139/ssrn.4487146>

- [28] Liu, L., Fan, K., Yang, M. (2023). Federated learning: A deep learning model based on resnet18 dual path for lung nodule detection. *Multimedia Tools and Applications*, 82(11): 17437-17450. <https://doi.org/10.1007/s11042-022-14107-0>
- [29] Nakrani, M.G., Sable, G.S., Shinde, U.B. (2020). ResNet based lung nodules detection from computed tomography images. *International Journal of Innovative Technology and Exploring Engineering*, 9(4): 1711-1714. <https://doi.org/10.35940/ijitee.D1540.029420>
- [30] Kumaran S.Y., Jeya, J.J., Khan, S.B., Alzahrani, S., Alojail, M. (2024). Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. *BMC Medical Imaging*, 24(1): 176. <https://doi.org/10.1186/s12880-024-01345-x>
- [31] Jin, H., Yu, C., Gong, Z., Zheng, R., Zhao, Y., Fu, Q. (2023). Machine learning techniques for pulmonary nodule computer-aided diagnosis using CT images: A systematic review. *Biomedical Signal Processing and Control*, 79: 104104. <https://doi.org/10.1016/j.bspc.2022.104104>
- [32] Nagaraja, P., Chennupati, S.K. (2023). Integration of adaptive segmentation with heuristic-aided novel ensemble-based deep learning model for lung cancer detection using CT images. *Intelligent Decision Technologies*, 17(4): 1135-1160. <https://doi.org/10.3233/IDT-230071>
- [33] Ali, I., Muzammil, M., Haq, I.U., Amir, M., Abdullah, S. (2021). Deep feature selection and decision level fusion for lungs nodule classification. *IEEE Access*, 9: 18962-18973. <https://doi.org/10.1109/ACCESS.2021.3054735>
- [34] Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J. (2016). RETAIN: Interpretable predictive model in healthcare using reverse time attention mechanism. *CoRR* abs/1608.05745. <https://doi.org/10.48550/arXiv.1608.05745>
- [35] 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>
- [36] 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>
- [37] Li, J., Gao, M., Wang, P., Li, B. (2024). Fall detection algorithm based on pyramid network and feature fusion. *Evolving Systems*, 15(5): 1957-1970. <https://doi.org/10.1007/s12530-024-09601-9>
- [38] Adesso, P., Restaino, R., Vivone, G. (2021). An improved version of the generalized Laplacian pyramid algorithm for pansharpening. *Remote Sensing*, 13(17): 3386. <https://doi.org/10.3390/rs13173386>
- [39] Wang, F., Silvestre, G., Curran, K.M. (2024). Segmenting fetal head with efficient fine-tuning strategies in low-resource settings: An empirical study with U-Net. *arXiv preprint arXiv:2407.20086*. <https://doi.org/10.48550/arXiv.2407.20086>
- [40] Rayachoti, E., Vedantham, R., Gundabatin, S.G. (2024). EU-net: An automated CNN based ebola U-net model for efficient medical image segmentation. *Multimedia Tools and Applications*, 83(30): 74323-74347. <https://doi.org/10.1007/s11042-024-18482-8>
- [41] Liu, H. (2024). Brain CT to MRI medical image transformation based on U-Net. *Journal of Physics: Conference Series*, 2824(1): 012002. <https://doi.org/10.1088/1742-6596/2824/1/012002>
- [42] Ju, M., Luo, J., Wang, Z., Luo, H. (2021). Adaptive feature fusion with attention mechanism for multi-scale target detection. *Neural Computing and Applications*, 33: 2769-2781. <https://doi.org/10.1007/s00521-020-05150-9>
- [43] Apostolidis, E., Balaouras, G., Mezaris, V., Patras, I. (2021). Combining global and local attention with positional encoding for video summarization. In *2021 IEEE International Symposium on Multimedia (ISM)*, Naples, Italy, pp. 226-234. <https://doi.org/10.1109/ISM52913.2021.00045>
- [44] Deng, L., Liu, B., Li, Z. (2024). Multimodal sentiment analysis based on a cross-modal multihead attention mechanism. *Computers, Materials & Continua*, 78(1): 1157-1170. <https://doi.org/10.32604/cmc.2023.042150>
- [45] Abbasi, M., Saeedi, P. (2024). Enhancing multivariate time series classifiers through self-attention and relative positioning infusion. *IEEE Access*, 12: 67273-67290. <https://doi.org/10.1109/ACCESS.2024.3397783>
- [46] Liu, C., Gu, P., Xiao, Z. (2022). Multiscale U-net with spatial positional attention for retinal vessel segmentation. *Journal of Healthcare Engineering*, 2022(1): 5188362. <https://doi.org/10.1155/2022/5188362>
- [47] Huang, X.Y. (2023). Research on medical image feature fusion and segmentation algorithm based on deep learning. *Changchun University of Technology*.
- [48] Kim, J., Lee, J. (2022). Global positional self-attention for skeleton-based action recognition. In *2022 26th International Conference on Pattern Recognition (ICPR)*, Montreal, QC, Canada, pp. 3355-3361. <https://doi.org/10.1109/ICPR56361.2022.9956156>
- [49] Tadepalli, Y., Kollati, M., Kuraparthi, S., Kora, P. (2021). EfficientNet-B0 based monocular dense-depth map estimation. *Traitement du Signal*, 38(5): 1485-1493. <https://doi.org/10.18280/ts.380524>
- [50] Cao, Z., Li, R., Yang, X., Fang, L., Li, Z., Li, J. (2023). Multi-scale detection of pulmonary nodules by integrating attention mechanism. *Scientific Reports*, 13(1): 5517. <https://doi.org/10.1038/s41598-023-32312-1>
- [51] Zhang, X., Liu, C., Yang, D., Song, T., Ye, Y., Li, K., Song, Y. (2023). RFACConv: Innovating spatial attention and standard convolutional operation. *arXiv preprint arXiv:2304.03198*. <https://doi.org/10.48550/arXiv.2304.03198>
- [52] Seung, H.K., Ho, C.K. (2023). A study on the classification of cancers with lung cancer pathological images using deep neural networks and self-attention structures. *Journal of Population Therapeutics and Clinical Pharmacology*, 30(6): 374-383. <https://doi.org/10.47750/jptcp.2023.30.06.042>
- [53] Chang, Y., Zheng, Z., Sun, Y., Zhao, M., Lu, Y., Zhang, Y. (2023). DPAFNet: A residual dual-path attention-fusion convolutional neural network for multimodal brain tumor segmentation. *Biomedical Signal Processing and Control*, 79: 104037. <https://doi.org/10.1016/j.bspc.2022.104037>

- [54] Ling, Z., Xin, Q., Lin, Y., Su, G., Shui, Z. (2024). Optimization of autonomous driving image detection based on RFACnv and triplet attention. *Applied and Computational Engineering*, 67: 68-75. <https://doi.org/10.54254/2755-2721/67/2024MA0067>
- [55] Hu, J., Peng, X., Xu, Z. (2012). Study of gray image pseudo-color processing algorithms. In 6th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Large Mirrors and Telescopes, 8415: 323-328. <https://doi.org/10.1117/12.977197>
- [56] 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>