VNU Journal of Science:
Computer Science and Communication Engineering
Journal homepage: http://www.jcsce.vnu.edu.vn/index.php/jcsce

Original Article

ResUNet Model Enhanced with Multiple Attention Mechanisms for Effective Pulmonary Nodule Segmentation in CT Images

Khai Dinh Lai[1,2,3], Thai Hoang Le[1,2*], Thuy Thanh Nguyen[4]

[1] *Faculty of Information Technology, University of Science, Nguyen Van Cu, Ho Chi Minh, Vietnam*
[2] *Vietnam National University Ho Chi Minh, Ho Chi Minh, Vietnam*
[3] *Saigon University, An Duong Vuong, Ho Chi Minh, Vietnam*
[4] *VNU University of Engineering and Technology, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam*

Received 16 August 2024
Revised 03 December 2024; Accepted 12 March 2025

Abstract: This research proposes the Position combined Channel attention module - Residual Unet model (PCAM-ResUnet), an enhanced ResUnet++, to improve CT lung nodule segmentation. In the paper, the Squeeze-and-Excitation Block is replaced by Channel Attention module (CAM) and Position Attention module (PAM) respectively. More importantly, these two modules are combined to create the Position combined Channel attention module (PCAM), a new breakthrough in our model structure. Through a multi-stage training process, PCAM-ResUnet was evaluated on a test dataset comprising 2000 pulmonary nodule samples. The PCAM variant demonstrated outstanding performance, achieving an average Dice Similarity Coefficient (DSC) of 85.96%. It achieved 'Excellent' segmentation results (cases with a DSC ≥ 80%) in 82.80% of cases, while reducing the 'Needs Improvement' level (DSC < 40%) to 1.85%. The obtained results emphasize the effectiveness of PCAM-ResUnet, affirming its superiority and showcasing its considerable potential for widespread clinical applications in the medical field.

Keywords: PCAM-ResUnet, Pulmonary Nodule Segmentation, CT Scanning, Multi-Stage Training, Attention Mechanism.

1. Introduction

Lung cancer, which poses a substantial risk of death and is notoriously difficult to detect in its early stages [1], is now better comprehended as a result of medical progress; this clarity enables us to proactively confront and implement efficacious treatments, thus increasing our optimism and capacity to combat the disease. Commencing with minute lesions

* Corresponding author.
E-mail address: lhthai@fit.hcmus.edu.vn

https://doi.org/10.25073/2588-1086/vnucsce.3423

referred to as lung nodules, which are less than 3 mm in diameter, encircled by lung tissue that is ambivalent in nature and especially challenging to discern when in proximity to or connected to blood vessels [2]. While the majority are benign, a few may be indicative of lung cancer in its nascent stages. Effective management of lung nodules is of utmost importance in order to rule out the possibility of malignancy, achieve prompt treatment, and minimize avoidable patient complications [3]. In conjunction with the patient's medical history, evaluating the location, size, and shape of pulmonary nodules assists in determining the most effective monitoring or treatment strategy, thereby increasing the likelihood of recovery and decreasing potential risks.

Diagnosis in modern medicine is predicated on clinical and subclinical symptoms; for this reason, diagnostic imaging systems such as X-rays, CT, and MRI have grown in importance. These systems provide accurate images by utilizing cutting-edge software and technology. These technologies are indispensable instruments for the detection of infections and pulmonary nodules, in addition to being non-invasive. Upon obtaining and diagnosing patients presenting with ambiguous respiratory symptoms, physicians initially utilize X-rays to obtain a comprehensive yet restricted level of detail. When X-rays detect abnormal signs, doctors often turn to CT or MRI for more detailed images. In clinical practice, MRI is less used for lungs due to low proton density and rapid signal decay from the sensitive magnetic field. The lung's air-rich structure impedes sharp imaging, reducing resolution and contrast, making it difficult to detect small abnormalities. Noise from natural movements such as breathing and heartbeat also degrades image quality [4]. CT uses X-rays to create detailed three-dimensional images of the body, with the air in the lungs providing natural contrast, helping to clearly detect abnormalities and information about tumors. The quick scanning process and short breath-hold time also contribute to clearer images [5].

Imaging tools such as CT scans offer vital information, but their interpretation demands careful analysis and expertise from clinicians. Computer-Aided Diagnosis systems aid physicians in assessing medical images, highlighting important characteristics and potential indicators of sickness to facilitate clinical decision-making. Lung computer-aided detection systems are developed through a multi-step process, with each stage serving an important function in creating a thorough diagnostic system [6, 7].

The lung analysis Computer-Aided Detection (CADe) system employs a comprehensive strategy (see figure 1), wherein each component is specifically designed to surmount unique challenges associated with the identification and classification of lung nodules on CT scans. The procedure commences with a phase of data collection, during which medical images are gathered from various sources, with a particular emphasis on CT imaging due to its ability to provide detailed insights into lung structures [8]. Following this, in the preprocessing stage, endeavors to reduce noise and improve the quality of the image are supplemented by accurate lung segmentation in computed tomography scans [9].

As the system progresses to the third phase, Nodule Detection, its objective is to precisely identify prospective sites of lung nodules. This process encompasses the identification of potential nodules as well as the allocation of probabilities to each, denoting the extent to which it is likely to be an authentic nodule.

The process culminates in the fourth stage, the False Positive Reduction stage aims to minimize the occurrence of false positives among the selected candidate sites. This stage involves a classification task that aims to differentiate between nodules and non-nodules.

As a result, it helps to simplify the list of lung nodules that need to be addressed. The CADe system's painstaking approach showcases its advanced ability to negotiate the intricate landscape of lung nodule identification and categorization, highlighting its substantial potential for clinical use [10].

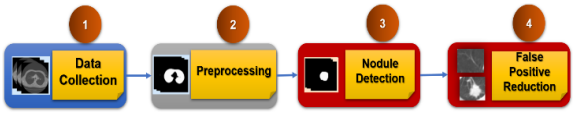

Figure 1. Four Stages of the CADe Process for Pulmonary Nodule Analysis.

This study largely concentrates on the third phase of the CADe system, which involves utilizing image segmentation to precisely outline pulmonary nodules. We aim to improve the accuracy in identifying and evaluating the likelihood of possible nodules, thus enhancing diagnostic efficiency and aiding in clinical decision-making.

Medical image segmentation is the process of breaking an image into discrete segments or areas, each representing a unique object or type of structure. Image segmentation in the context of CT lung nodules involves accurately identifying the specific locations of lung nodules in computed tomography imaging of the lungs. A machine learning model with the ability to distinguish between lung nodules and the surrounding tissue, as well as adjacent features such as blood arteries and the lung itself, is necessary for this task. The difficulties in precisely dividing lung nodules mostly arise from their diminutive dimensions, particularly when they are situated at the periphery of the lungs or in close proximity to blood veins. The method of segmenting lung nodules involves a wide variety of architectural designs, image preprocessing techniques, and training strategies [11]. Several lung nodule segmentation approaches based on deep learning employ multi-view neural network architectures, while others utilize generic neural network structures. The method employs a multi-view neural network design to incorporate many viewpoints of the lung nodule and utilize them as input for the neural network. Meanwhile, the generic neural network architecture is constructed by modifying or incorporating additional blocks into existing CNNs.

The main contribution of this work resides in:

(1) The innovative Position combined Channel Attention Module (PCAM) is the result of combining the Channel Attention Module (CAM) and the Position Attention Module (PAM), both of which are well-established attention mechanisms. While PAM is refined to analyze and generate spatial features from the relationships between pixels, CAM is adjusted to exploit the relationships between feature channels, thereby capturing important information from various perspectives of the image. The innovation of PCAM not only comes from integrating information from both spatial and channel dimensions but also from how these two modules are customized to work together effectively, bringing benefits from both attention mechanisms for more accurate medical image segmentation.

(2) PCAM-ResUNet is an improved iteration of the ResUnet++ architecture that is being proposed. The primary enhancement of this model is the incorporation of the PCAM, which functions as a substitute for the SE blocks that were previously present in the ResUnet++ architecture. We conducted an experiment in which we substituted the CAM, PAM, and PCAM modules for the SE Block. The resulting three primary variants of the model were designated with the names of the replacement modules: CAM focuses on channel-specific features, PAM targets spatial features, and PCAM is the complete combination of both types of attention.

(3) Experiments conducted on the LUNA16 dataset showed notable enhancements exhibited by the novel model in comparison to the initial version. These results confirm the potential and efficacy of PCAM-ResUNet in facilitating fresh possibilities for scientific investigation and practical implementation in the domain of medical image segmentation.

Some techniques employed to enhance precision encompassed the following:

(1) The Generalized Center-Based Image Cropping (BCBIC) algorithm is introduced to enrich data by utilizing a shift factor generating technique. This technique facilitates the creation of several new images containing lung nodules, derived from the original image.

(2) The two-stage training approach is suggested, with the initial stage focused on acquiring knowledge of the overall characteristics of CT lung images as well as the diverse anomalies present in different areas. In the second stage, the complex shapes of lung nodules are studied in depth.

The paper is methodically structured into sections to offer a coherent and comprehensive examination of the topic matter. The literature review on lung nodule segmentation in CT scans is included in Section 2, where the essay discusses typical attention mechanisms and neural network integration methodologies. This section also explores the history of Unet, the U-shaped model, and its improvements, with a focus on attention mechanisms. Section 3 details the core parts of our method. This section examines the channel and position attention modules, which led to the PCAM. The essential components and their roles in the proposed model will be explained. Section 4 discusses 'Experiment Results and Analysis'. We start with the LUNA16 public database, which helps us validate our experiments. Datasets for testing and assessment and data augmentation strategies to improve model resilience will be discussed next. This study stands out for comparing segmentation results from PCAM-ResUnet and ResUnet++ models. Segmentation results by nodule size and overall are compared to assist choose the optimum model configuration. Benchmarking against SOTA techniques is undertaken. Section 5, 'Discussion', discusses our model's improvements and evaluates its learned parameter values.

2. Literature review

2.1. Contemporary Models for Segmenting Lung nodules in CT scans

Studies have demonstrated that CNN topologies can improve the effectiveness of lung segmentation methods [12-14] . Out of them, segmentation networks such as Fully Convolutional Neural Network (FCN) [15] and U-Net [16] have been highly acclaimed.

Expanding upon these networks, some segmentation studies have refined and adapted their models by utilizing the fundamental CNN design or by modifying or incorporating additional components into the existing CNN structure. As an illustration, Huang et al. [17] introduced a system comprising four primary modules: utilizing a Faster regional-CNN (R-CNN) to discover nodule candidates, consolidating the candidates, employing a CNN to decrease false positives, and segmenting nodules using a customized FCN. Their model underwent training and validation using the LIDC-IDRI dataset, resulting in an average Dice Similarity Coefficient (DSC) of 0.793. Tong et al. [18] utilized the U-Net architecture to perform lung nodule segmentation. Their approach improved network performance by combining U-Net with a residual block. In addition, the lung field was isolated using morphological techniques, and the images were resized to a dimension of 64 × 64 pixels before being fed into their network. The model under consideration conducted training and validation using the LUNA16 dataset, resulting in a DSC of

0.736. Keetha et al. [13] introduced a resource-efficient U-Det architecture by merging U-Net with Bi-FPN, which is implemented in Efficient-Det. The network was trained and tested on the LUNA dataset, achieving an average Dice Similarity Coefficient (DSC) of 82.82% and an average Sensitivity (SEN) of 92.25%.

2.2. Attention Mechanism

Within the human visual perception system, we possess an innate capacity to concentrate on significant regions and disregard secondary or irrelevant data in the surroundings. This enhances our aptitude to precisely and effectively distinguish and categorize stimuli [19]. By imitating this capacity, the attention mechanism was established to provide a weight-based approach for adjusting the focus to various parts within an image. This enables neural networks to selectively concentrate on relevant regions that are connected to the purpose while disregarding irrelevant ones. The absorption power of this method is exceptionally efficient in capturing intricate semantic links in the segmentation of medical images. Moreover, the attention mechanism is valuable for elucidating the correlation between input and output data, facilitating the visualization of the model's acquired knowledge, thus offering a perceptive and clear understanding of the intricate structure of neural networks.

2.2.1. Channel Attention

Channel attention refers to the process of adjusting to individual data channels by considering them as distinct representations of various objects, as proposed by Chen et al. (2017) [37]. The concept of the squeeze-and-excitation network (SENet) was initially introduced by Hu [20] and his research team. The squeeze module compresses each channel into a singular value via global average pooling, while the exciting module generates an attention vector using fully connected and nonlinear layers.

Subsequent endeavors have sought to enhance the compression or activation procedures. In their study, Qin et al. [21] considered global average pooling as a specific instance of the discrete cosine transform in the squeeze module. They then introduced the Frequency Channel Attention Network (FcaNet), which is designed to compress information. Wang et al. [22] developed an Efficient Channel Attention (ECA) block that utilizes direct interactions between each channel and its k-nearest neighbors to improve the excitation module with reduced complexity. Lee et al. [23] used style pooling in the squeezing phase and added a fully connected per-channel layer in the excitation module to decrease computing expenses in both stages.

2.2.2. Spatial Attention

In the realm of image analysis, we employ spatial attention approaches to identify significant areas within the image. This is accomplished by assigning scores to separate spatial regions on the feature map, which are calculated based on their width and height.

Oktay et al. [24] state that the attention gate employs a cumulative attention mechanism to generate gating coefficients by combining the input and gate signal on a global level. The outcome is a weight map that focuses on spatial attention, resulting in a model that is both efficient and impactful by highlighting significant regions and disregarding irrelevant characteristics. The GENet, as proposed by Hu et al. [25], incorporated a spatial recalibration function called the gather-excite module.

2.2.3. The Combination and Position of Attention Modules

The attention mechanism is commonly incorporated into deep learning models as an additional layer, similar to a "plugin" that may be placed at any point inside the network's convolutional blocks. The placement of this insertion is flexible and is determined by the model's needs and the task's features. The attention mechanism is commonly used in three key areas: the encoder, to help the model focus on crucial input features; the decoder, to improve output reconstruction by focusing on important

encoded information; and skip connections, to prevent loss of detailed information across layers. A hybrid approach can be used by placing the attention mechanism at different positions to gather information from both channels and spatial features effectively. This enhances the model's capability to identify features and improves the accuracy of segmentation maps or classification.

The Convolutional Block Attention Module (CBAM), introduced by Woo et al. [26], computes channel attention and spatial attention in a sequential and independent manner. The channel attention module utilizes two parallel branches with max-pooling and avg-pooling operations, whereas the spatial attention module employs a convolutional layer with a bigger kernel to create the attention map. CBAM can prioritize effective routes and strengthen certain areas with crucial information.

The Dual Attention Network (DANet), introduced by Fu et al. [27], utilizes self-attention to independently calculate channel attention and spatial attention, which are then combined to provide the ultimate outcome. DANet collects data from channels and space separately and then combines it to build a thorough comprehension of the incoming data's structure and features.

2.3. U-shaped model Incorporating an Attention Mechanism

Long et al. [15] introduced the "skip" architecture in fully convolutional networks, a significant advancement in image segmentation by enabling precise division of images without complex post-processing. This approach enhances spatial detail recovery lost in downsampling and laid the groundwork for models like U-Net [16] by Ronneberger, Fischer, and Brox. U-Net’s structure includes a contracting path for feature extraction and an expanding path with skip connections that merge multi-level feature information, enhancing pixel-level segmentation accuracy. Optimized for medical applications, U-Net performs well with limited data. However, it faces challenges in multi-task, complex scenarios and can overfit with limited data, performing poorly on new data. It also requires substantial processing power, especially for high-resolution images, limiting its use in resource-constrained environments. Consequently, it has been enhanced and broadened through several iterations as [28, 29]. The versions aim to improve feature learning, reduce overfitting, and optimize segmentation. ResUnet, an enhanced version of U-Net, was developed by Zhang et al. [30] to overcome U-Net's constraints by incorporating residual connections inspired by He et al.'s ResNet model [31]. This enhancement enhances the deep learning capability and tackles multi-task challenges without significantly raising computing resource requirements, rendering ResUnet more appropriate for contexts with limited resources. B. John Jaidhan and Banavathu Sridevi [32] created DAH-UNet, a modified UNet architecture that combines residual blocks, enhanced atrous spatial pyramid pooling (ASPP), and depth-wise convolutions. This adaptation, together with a boundary-aware hybrid loss function, has demonstrated higher accuracy on two public datasets than current models.

However, low resolution and contrast in pictures, such as in lung CT scans, can reduce the ability to differentiate intricate and comparable structures. To precisely segment abnormal structures like lung nodules or lesions, a model must be able to identify and distinguish delicate characteristics, a challenge that Residual layers may not always effectively address. ResUNet++ was created by Debesh Jha et al. [33] to tackle these issues with notable enhancements. The Squeeze-and-Excitation block (SE block) is a significant improvement that assists the model in concentrating on distinctive features, hence enhancing accuracy and removing unnecessary information. ResUNet++ utilizes Atrous Spatial Pyramid Pooling (ASPP) [34] to gather input from different spatial scales, helping the model comprehend and analyze structures of varying

sizes more effectively. Additionally, the decoder section of ResUNet++ is improved with attention modules to help the model concentrate on important regions of the image when reconstructing the segmentation output. The modules allow the model to adapt its focus dynamically, enhancing the detection and segmentation of items of interest, leading to improved accuracy and efficiency in the segmentation process. Figure depicts the transition from Unet to ResUnet++.

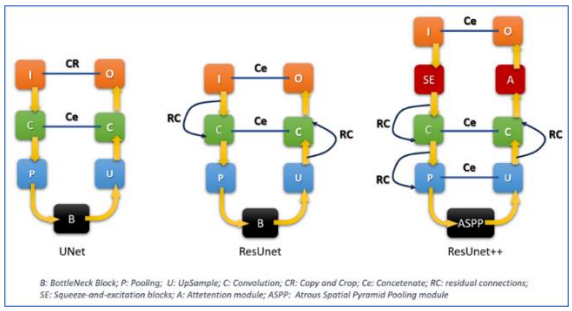

Figure 2. An overview of the distinctions among the Unet, ResUnet, and ResUnet++ models.

SE blocks, utilizing channel filters, do not use the interdependence of local image attributes. Hence, more advancements appropriate for segmentation tasks are required. Our research aims to expand and improve the model, using the specific details provided in Section 3: Proposed Model.

3. Architecture and Components of the Proposed Model

The model we propose is derived from the ResUnet++ model but substitutes the SE block with a combination of two modules: channel attention and position attention. This section will outline the techniques of the key modules and emphasize their functions in the model.

We will now explore the attention mechanism in the algorithm, with a specific focus on Computing Attention, Feature Refinement Using Attention, and Adjusting the Impact of Attention.

3.1. Channel Attention Module (CAM)

Deeply comprehending and effectively utilizing image feature channels are crucial for improving the performance of deep learning models in challenging tasks within computer vision. The Channel Attention Module (CAM) enhances the process by refining feature channels to emphasize critical qualities and eliminate irrelevant information. We present the CAM algorithm (see in figure 3,4,5) and then explore the core of the design. We will examine the operating principles and evaluate the influence of CAM on the efficiency of the proposed model.

Algorithm: Channel Attention Module (CAM)
Input:
- X: Input feature map with dimensions (batch_size (B), channels (C), height (H), width (W)).

Output:
- O: Output feature map with channel-wise enhanced features.

Procedure:
Step 1: Feature Map Reshaping
- $A = reshape(X,[B, C, H \times W])$
- $A^T = reshape(X,(B, H \times W, C))$

Step 2: Compute Attention Map
- $Ma = A \times A^T$

Step 3: Normalize Attention Map
Initialize attention_batch_new with the same dimensions as attention_batch
- For each batch_index b in 0 to B-1:
- attention_new = *NormalizeA*(attention_batch[b])
- attention_batch_new[b] = attention_new
- M_c = Softmax (attention_batch_new)

Step 4: Apply Attention to Features
- $F = M_c \times A$
- $Ms = F.ReshapeTensor(B, C, H, W)$

Step 5: Combine Attention Features with Input:
- $O = \beta * Ms + X$

End of Procedure

Figure 3. Pseudocode for the CAM algorithm.

```
Function NormalizeA (attention):
  // attention is a matrix with dimensions (H, W)
  Initialize attention_new with the same dimensions as attention
  Initialize max_values with dimensions (H, 1)

  // Find the max value for each row
  For each row index h in 0 to H-1:
    max_value = -Infinity
    For each col index w in 0 to W-1:
      If attention[h, w] > max_value:
        max_value = attention[h, w]
    max_values[h] = max_value

  // Expand max_values to the same size as attention and subtract
  For each row index h in 0 to H-1:
    For each col_index w in 0 to W-1:
      expanded_max = max_values[h]
      attention_new[h, w] = expanded_max - attention[h, w]

  Return attention_new
```

Figure 4. Pseudocode for the NormalizeA algorithm.

We simplify the algorithm's function within the model by assuming that the batch size (*B*) is 1, as depicted in Figure 6. This method aids in elucidating each operation and its impact within the Channel Attention Module (CAM), facilitating the comprehension of the fundamental operational process.

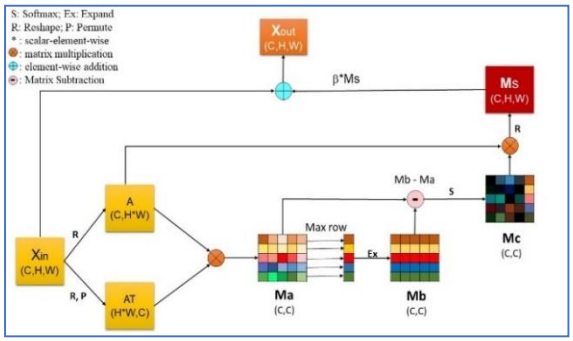

Figure 5. Channel Attention Module algorithm Architecture.

a. Attention Calculation Method:

This algorithm segment is dedicated to evaluating the significance of each feature in the input data. This approach entails utilizing operations like the dot product or comparable mechanisms to quantify the association among characteristics, resulting in the creation of an attention map. The attention map illustrates the significance of each piece and is created during the initial three steps of the algorithm to clearly pinpoint and highlight the regions that need concentration in further processing.

+ Step 1: Feature Map Reshaping

In this step, the input feature map *X*, which is three-dimensional according to batch size, number of channels, height, and width, is reshaped to prepare for attention computation:

• Tensor *X* is reshaped into matrix *A* with dimensions *[B, C, H×W]*, converting the three-dimensional tensor into a two-dimensional matrix where each row corresponds to a distinct vector for each channel in each batch.

• Tensor *X* is reshaped into *AT*, with channels and spatial dimensions interchanged, in preparation for matrix multiplication with *A*.

+ Step 2: Compute Attention Map

At this step, the attention matrix is computed.

• $Ma = A \times AT$: Matrix multiplication between *A* and *AT* is performed to generate the attention matrix *Ma*. This multiplication is not element-wise but matrix multiplication, where each element of Ma represents the degree of correlation between feature channels for each batch. The result of this multiplication yields a matrix containing the correlation values between features across the entire feature space.

+ Step 3: Normalize Attention map

Each attention matrix *Ma* in the tensor is analyzed individually, corresponding to each sample in the batch. We identify the highest value in each row of the *Ma* matrix to generate the *Mb* matrix, where each element is the maximum value from the respective row of *Ma*. The tensor *Mb*, which holds the maximum values, is then enlarged to match the size of *Ma*. Subtracting Ma from *Mb* yields the difference tensor *Mc*. Each value in *Ma* is compared to the maximum value in its respective row, creating a difference matrix where the maximum value in each row is normalized to 0, and other values represent the deviation from that maximum. The softmax function is used on *Mc* to transform the difference matrix into a normalized attention matrix for each sample in the batch. Within this normalized attention matrix, a value of 0 signifies the utmost attention given to the original maximum value, with other values indicating varying degrees of attention based on their deviation from the maximum. This technique improves focus on key features in each batch sample and allows the model to evaluate and alter attention weights flexibly, enhancing the performance of processing multidimensional data.

b. Feature refinement is enhanced by employing attention up to the fourth step of the algorithm.

+ Step 4: Apply Attention to Features

Once the attention map Mc is computed, feature refinement takes place as follows:

• *Attention Weighting*: The attention map *Mc* is normalized using the softmax function to create a probability distribution, which is then used to weight the feature map *A* using matrix multiplication. The product of this multiplication is denoted as $F=Mc \times A$, where each element of A is multiplied by its associated coefficient from Mc. This technique refines features by assigning larger weights to more essential feature channels learned through the attention process, and de-emphasizing less significant ones.

• *Feature Enhancement*: This procedure identifies significant feature channels and adjusts the feature response levels according to the model's observations from the data. Refined F contains detailed feature information ideal for picture segmentation, emphasizing crucial areas.

• *Spatial Consistency*: An important point to note is that this process maintains the spatial structure of the original feature map. Although each feature is weighted differently, the overall space it occupies is preserved. This ensures that the detailed spatial information necessary for medical image segmentation is not lost during the refinement process.

c. Adjusting the Impact of Attention is carried out through step 5 Combine Attention Features with Input in the algorithm.

• *Feature Integration*: The enhanced attention feature map *Mc* is integrated with the original feature map *X* via a weighted addition, controlled by a trainable scale factor β. This facilitates the incorporation of data from Mc and *X* and also empowers the model to autonomously regulate the influence of the attention process according to previous computations, resulting in an output $O = \beta \cdot Mc + X$ with improved characteristics.

3.2. Position Attention Module (PAM)

The Position Attention Module is a spatial attention mechanism that aims to improve the model's sensitivity to locations that are likely to have valuable information about lung nodules. Examining spatial concentration levels for each position in the image is a crucial element of the Position Attention Module. The procedure provides detailed and optimized spatial information in a refined manner as figure 7 and figure 8:

Algorithm: Position Attention module

Input:
- *I*: Input image.
- *C*: Number of image channels.

Output:
- *O*: Output image.

Procedure:

Step 1: Size normalization
- $Idown = DownSample(I)$

Step 2: Feature extraction
- $Fb = Conv(Idown, C/8)$
- $Fc = Conv(Idown, C/8)$
- $Fd = Conv(Idown, C)$

Step 3: Attention matrix construction
- $As = Softmax(FbF_c^T)$

Step 4: Refinement of features using attention
- $Fe = Fd\ A_s^T$

Step 5: Size restoration of feature representation
- $Fe = Interpolate(Fe, size = (Hup, Wup))$
- $Iup = Interpolate(Idown, size = (Hup, Wup))$

Step 6: Combining features and attention
- $O = \alpha Fe + Iup$

Where:
- *DownSample*: Function to reduce spatial size of image.
- *Conv*: Convolution function with specified output channel size.
- *Softmax*: Probability normalization function.
- *Interpolate*: Function to resize the image to new dimensions.
- *T*: Tensor transpose.
- α: Learned scaling coefficient.
- *Hup*,*Wup*: Target height and width for upsampling.

End Procedure

Figure 6. Pseudocode for the NormalizeAttention algorithm.

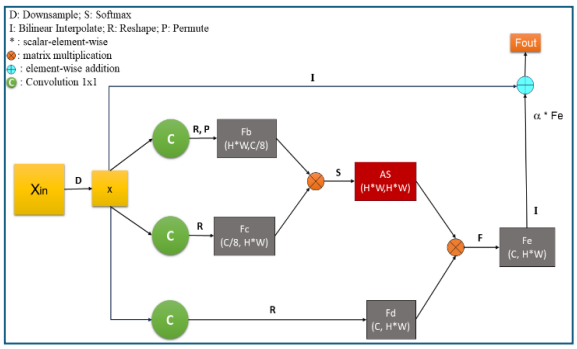

Figure 7. Position Attention Module Architecture.

a. Attention Calculation Method:

+ *Step 1 Size normalization*: Initially, the input image I is spatially downsampled through the DownSample function to reduce computational complexity and focus on more significant information at a higher level. This reduced sample size is typically much smaller than the original image, reducing the spatial dimensions to be processed.

+ *Step 2 Feature Extraction*: The input Xdown is processed through convolutional layers to extract spatial feature maps. Fb, Fc are extracted via a convolutional layer with the output channel number being C/8. This layer detects lower-level features that may include information about edges, corners, and other basic structural characteristics of the image. Fd is extracted with the full channel depth C, retaining more diverse and complex feature information from the downsampled image. Each convolutional layer is usually paired with normalization methods such as batch normalization and activation functions like ReLU to enhance the learning process and add non-linearity to the model. Convolutional layers are typically set up with specific strides and padding to regulate the size of the feature map output and prevent the loss of important spatial features during extraction.

+ *Step 3 Attention matrix construction*: The matrix *Fb* is multiplied by the transpose of matrix *Fc* to generate a spatial correlation matrix, which indicates the correlation strength between each pair of places in the feature space. The softmax function is used to normalize each row of the matrix into a probability distribution after the multiplication, in order to emphasize spots with high correlation levels.

The purpose of the Attention Matrix is to focus on specific elements. The softmax normalization step is designed to adapt spatial features by assigning a weight to each position according to its relationship with other positions. Within medical imaging, this allows the model to identify crucial locations requiring attention, like lung nodules, by concentrating on regions with strong correlation in the feature space. The attention matrix is more than just a weight map; it is the outcome of a sophisticated calculation that models spatial correlation. It plays a vital role in enhancing the accuracy of medical picture segmentation tasks.

b. Enhancing features with Attention mechanism

Following the attention matrix Feature refining is conducted to improve the quality of feature information by using the acquired spatial correlations. This phase takes place at step 4 of the algorithm using the following techniques:

• *Application of Attention Matrix*: The feature map *Fd*, derived from the last convolutional layer with the complete channel size to preserve varied information, undergoes matrix multiplication with the transposed attention matrix *As*. This multiplication involves matrix multiplication, where each element of the feature map is adjusted according to the relevant attention weights. This function either increases or decreases the feature signal at each point according to the spatial significance identified by the attention matrix.

• *Feature Refinement*: The outcome is a revised feature map *Fe*, where each feature has been modified to represent its significance based on both content and spatial aspects. For instance, in lung nodule segmentation, features at the nodules' location will be highlighted, whereas features in the background or unnecessary areas may be reduced.

• Preserving Spatial Information: It is crucial that this enhancement does not compromise the initial spatial information of the feature map, but instead enhances it. This guarantees that the intricate spatial data required for medical image segmentation is maintained.

c. Adjusting the impact of Attention

In the final part of the Position Attention Module, we adjust the impact of attention on the

final feature through a technical process involving steps 5 and 6 of the algorithm.

+ *Step 5 Size restoration*: Resize the original input Idown using the Interpolate function to match the scale of the attention feature map Fe before adding them together. This maintains spatial information and guarantees that the enhanced features are consistently implemented throughout the image.

+ *Step 6 combine features and attention*: The end outcome is O, achieved by meticulously merging the enhanced feature information from attention with the original input, following the described method.

• *Combining Feature Maps*: The outcome of the feature refinement process *Fe*, in which each feature is modified based on the attention matrix, is subsequently merged with the downsampled original input *Xdown*. Typically, this process involves merging information from *Fe* and *Xup* through element-wise addition after resizing the original input using interpolation.

• *Scaling Coefficient Acquired*: In this process, a scaling coefficient α is utilized to balance the impact of enhanced attention with the original information. The coefficient is acquired through the network training process and enables the model to autonomously regulate the blend of attention-based information with the original data.

3.3. Position combined Channel attention module (PCAM)

We suggest a novel attention module that incorporates a multi-Attention technique to enhance feature representation capabilities (refer to the figure 9, 10). The combination of CAM and PAM was chosen to leverage the strengths of each module: CAM enhances the ability to emphasize critical features along the channel dimension, while PAM improves focus on important spatial regions. This integration creates a robust attention map, optimizing both spatial and channel dimensions, thereby enabling the model to achieve higher performance in segmenting complex pulmonary nodules. The model consists of three primary steps, demonstrating the incorporation of theoretical concepts with practical application.

Procedure:
- **Step 1: Initialize Modules:**
  - Initialize spatial attention (**PAM**) and channel attention (**CAM**) modules.
  - Define convolution layer for feature refinement.
- **Step 2: Apply Dual Attention:**
  - Apply position attention to **x**: **p = self.position(x)**.
  - Apply channel attention to **x**: **c = self.channel(x)**.
- **Step 3: Combine Attention Outputs and Refine:**
  - Combine the outputs from position and channel attention: **combined = p + c**.
  - Refine the combined features using the convolutional layer: **enhanced_features = self.conv(combined)**.
- **Step 4: Return Output:**
  - Return the refined feature map **enhanced_features**.

End of Procedure

Figure 8. Pseudocode for the PCAM algorithm.

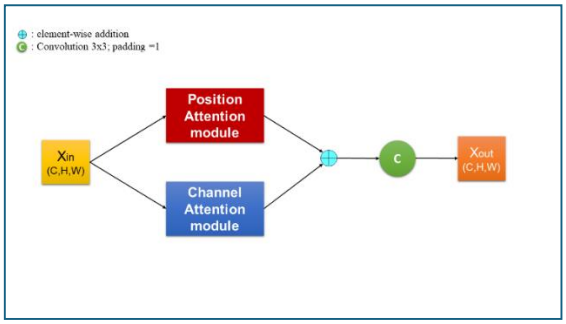

Figure 9. PCAM Architecture.

a. Attention Calculation Method and Feature Refinement

+ Step 1 Module Initialization

It involves setting up the Position Attention Module (PAM) and the Channel Attention Module (CAM) to utilize feature information from spatial and channel dimensions. This stage establishes the foundation for concentrating on specific regions of significance within the image.

+ Step 2 Compute the attention modules separately.

Spatial attention mechanism is utilized via PAM to generate a spatial attention map, highlighting the spatial connections among pixels. CAM is utilized to compute a channel attention map, which emphasizes the significance of relationships between channels.

+ Step 3 Feature combination and refinement

The results from PAM and CAM are merged through element-wise addition to form the

combined feature. A convolutional layer is employed to enhance and process the data, resulting in the final feature map.

b. Adjusting the impact of Attention

While training, the parameters α and β are modified to specify the influence levels of PAM and CAM on the combined feature map. The parameters are fine-tuned according to the model's error rate on the training set, enabling the model to adapt and concentrate on the crucial aspects of the image.

3.4. Proposed model

We introduce a meticulously constructed deep neural network model for accurately segmenting lung nodules from CT scans by effectively handling image characteristics. The model is built using a blend of sophisticated algorithms and meticulous data processing strategies. An extensive examination of each component of the model is provided in figure 11.

• *Encoder and Residual Blocks*:

The CT images are initially fed into the model using encoder blocks. The Residual Block (ResBlock) in this context integrates convolutional layers, batch normalization, and the ReLU activation function to extract and enhance low-level information from the image. The convolutional layer aims to retain crucial spatial characteristics, whereas residual connections aid in reducing the issue of disappearing gradients. This guarantees the preservation of spatial information and the inherent structure of the lungs, which is essential for future segmentation procedures.

• *PCAM*: The image is processed by the encoder blocks and then sent to the PCAM, which consists of the Channel Attention Module (CAM) and the Position Attention Module (PAM). CAM utilizes a self-attention process to examine the relationship between channels and concentrate on the channel that holds the most precise information regarding the lung nodule. PAM emphasizes spatial elements and highlights significant areas within the image. PCAM is a fusion of CAM and PAM, which generates a detailed spatial-channel attention map to emphasize significant regions and channels that hold lung nodule data.

• *ASPP Bridge và Residual Convolution:*

The Residual Convolution Modules improve feature representation by combining information from previous and current layers, optimizing detail and context. The Atrous Spatial Pyramid Pooling (ASPP) module expands the receptive field using Atrous Convolution, capturing a broader area of the input image by adjusting dilation rates without increasing parameters. This enhances the model's ability to understand both contextual and specific object details, critical for tasks like medical image segmentation.

In traditional convolution, the value of a point in the output feature map (activation map) is calculated by applying a filter (kernel) F to the input feature map I as follows:

$$O(x,y) = (I * F)(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x+i, y+j).F(x,y) \quad (1)$$

where $*$ denotes the convolution operation, O(x,y) is the value at point (x,y) on the output feature map $I(x+i,y+j)$ is the corresponding value on the input feature map, and $F(i,j)$ is the weight of the filter at position (i,j). In the convolution formula, the filter F has a size of $(2k+1)\times(2k+1)$, where k is defined as the 'radius' of the filter, not the full size. This allows the filter to cover a specific neighborhood on the input feature map to calculate the value for each point on the output feature map. (for example, for a 3x3 filter, k=1). Atrous convolution extends the above formula by adding a dilation parameter d, allowing for larger distances between weights in the filter:

$$O(x,y) = (I *_d dF)(x,y) \quad (2)$$
$$= \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x + d.i, y + d.j).F(x,y)$$

Here, $*d$ denotes the atrous convolution operation, and d is the dilation rate, which defines the distance between elements in the kernel applied to the input. When d=1, atrous convolution becomes traditional convolution.

The Residual Convolution Modules and ASPP Bridge process feature maps after PCAM. ASPP filters spatial information at many scales. Atrous convolution layers with variable dilation rates allow the model to evaluate vast receptive fields and capture lung nodule details. This information combination helps the model optimize segmentation, especially for lung nodules with diverse sizes, forms, or structures. The model can recognize lung nodules of various sizes and shapes using this method, enhancing segmentation. Concatenating feature maps from each ASPP block follows processing. A comprehensive feature collection with information from several spatial scales results from this combination. This technique is crucial for synthesizing varied information to show the item and its context in the image.

• *Attention Module:*

The Attention module is an important part that helps the model focus on the important parts of the image by using feature information from ASPP and the encoder layers. The feature from ASPP, which contains information at various spatial scales, is combined with information from the encoder layers to create a complete picture of the important features that the model needs to focus on. This process is carried out through a series of convolutional layers and ReLU activation functions to effectively process and combine this information.

The calculation formula in the Attention module can be represented as follows:

InterE(x1) = ReLU(BN(Conv(*x*1))) (3)

ConvE(*x*1) = MaxPool(Conv(InterE(x1))) (4)

ConvD(*x*2) = Conv(ReLU(BN(Conv (*x*2)))) (5)

Output = Conv (ConvE (*x*1) + ConvD(*x*2))·*x*2 (6)

MaxPooling reduces feature spatial extent after ConvE's initial convolutional layer. The formula matches the Attention module's sequence of operations. Applying Convolutional layers and MaxPooling to the encoder feature, Batch Normalization, and ReLU to the decoder feature are the next stages. It ensures that the Attention module may efficiently alter the decoder's feature weight based on the attention procedure's relevance.

• *Decoder và UpSampling:*

The model uses UpSampling and ConvTranspose2d layers to increase the feature space and recreate the image's fine resolution during decoding. This process is aided by skip connections from the encoder, which combine specific spatial information with the decoded characteristics to efficiently integrate both contextual and spatial detail information. Attention modules are reinstated at this point to highlight important sections and improve the segmentation output.

• *Final Segmentation Result:*

A secondary ASPP module adds spatial information processing after the decoder integrates and enhances data from Residual Convolution, PCAM, ASSP bridge, Attention module, and UpSampling. Integrating ASPP toward the end of the model, before using the sigmoid function, improves spatial data collection. This ensures that the final segmentation map accurately depicts lung nodule location, anatomy, and size. After the sigmoid layer, the model creates the lung nodule segmentation map, categorizing pixels by their chance of being lung nodules. The CT scan lung nodule map shows their location and structure.

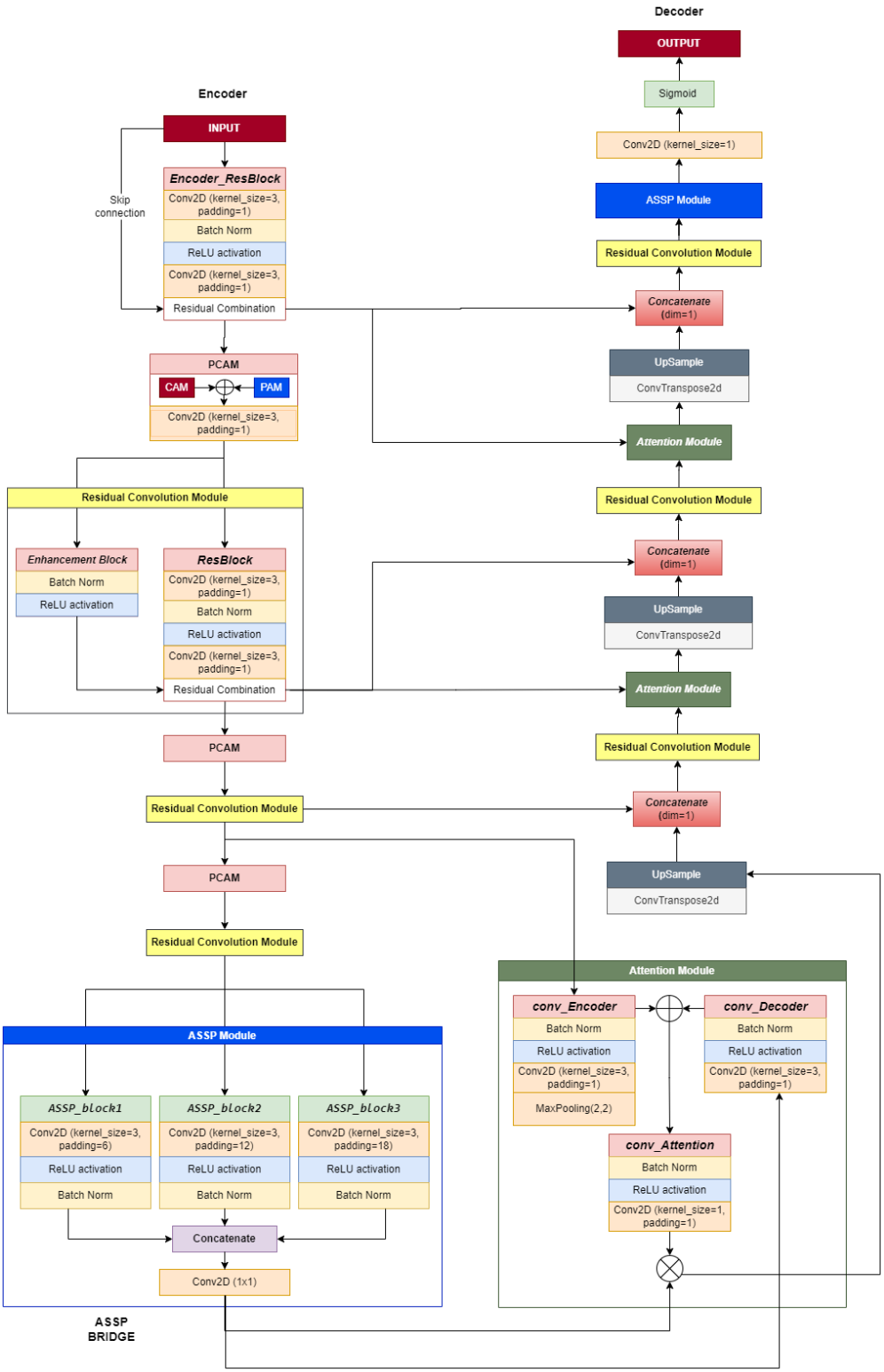

Figure 10. PCAM-ResUnet architecture.

4. Experiment result and analysis

4.1. Datatsets

4.1.1. LUNA16

The LUNA16 dataset, a component of the LUNA (LUng Nodule analysis) challenge, is notable in the realm of medical machine learning and CT lung image analysis. LUNA16 is released under the Creative Commons Attribution 4.0 International License, making it transparent and widely accessible, using the public LIDC/IDRI database. This dataset provides a substantial amount of top-notch CT images along with thorough annotations from skilled diagnostic radiologists (Setio et al., [35]).

LUNA16 chooses CT scans from the database based on a criterion that the slice thickness is no more than 2.5 mm, specifically focused on 888 high-quality CT scans. The lung nodules were marked and classified in the database by a thorough evaluation process including four competent diagnostic radiologists. Each radiologist participated in identifying and categorizing lung nodules to create a reference standard for the challenge. This standard included lung nodules that were 3 mm or larger and were agreed upon by at least three of the four radiologists.

The images are saved in the MetaImage format (mhd/raw), where each .mhd file is paired with a corresponding .raw file that holds pixel data. The annotation data is stored in a CSV file, providing information on the position and dimensions of each lung nodule. There are 1186 lung nodules documented in the annotation file.

LUNA16 is a dataset that serves as a significant challenge for the development and assessment of novel strategies in the detection and analysis of lung nodules. This dataset encourages researchers to compare the efficacy of automated algorithms, fostering advancements in medical imaging diagnoses.

4.1.2. Practical 512x512 dataset

We utilized 512x512 original size images, consisting of 1186 slice images with lung nodules retrieved from the LUNA16 annotation file. This dataset will be used for the early stage of training the models.

4.1.3. Practical 64x64 dataset

a. Set of data image

We cropped CT scans from the LUNA16 dataset, which were originally 512x512 pixels, into smaller segments of 64x64 pixels. This was done based on the center coordinates of lung nodules found in the annotation file. This decision was based on multiple explicit scientific and technological rationales. Cropping the image from 512x512 to 64x64 allows for a more focused view of the lung nodule and its surroundings by removing extraneous image components. This approach improves precision and effectiveness in examining crucial characteristics of lung nodules. The 64x64 pixel images strike a mix between image detail and quick processing, which optimizes model training and testing. Choosing image cropping using exact center coordinates guarantees that each image segment consistently includes the lung nodule and offers the required detailed information for precise analysis.

b. Data Augmentation

The dataset of images with lung nodules in the annotation is helpful but requires enrichment to align with the needs of deep learning models, particularly in scenarios that demand training with extensive and varied datasets to enhance accuracy and generalization.

We introduced the "Generalized Center-Based Image Cropping" algorithm to address this requirement (refer to the figure 12). This approach centers on utilizing the midpoint of the lung nodule, identified via annotations, as the reference point for cropping images. We change the center point of each lung nodule in the original 512x512 pixel image to crop multiple 64x64 pixel frames. This method guarantees that every cropped image includes the lung nodule and displays the variety of the surrounding area. In this study, although it was possible to get multiple images from each nodule, we decided

to generate 10 images for each original image. Consequently, we create 10 cropped images measuring 64x64 pixels from each original image. These cropped images depict the lung nodule from various orientations and perspectives, thereby enriching the diversity of the training data.

Algorithm: Generalized Center-Based Image Cropping

+ *Input:*

• Image (*img*): The initial image that will undergo cropping. This is the fundamental data that the algorithm functions on.

• Annotated Nodule Center Coordinates (Cx, Cy): The *x* and *y* coordinates denote the central point of the pulmonary nodule based on the data annotations. The image cropping will be determined by these coordinates to keep the nodule within the cropped image area.

• Diameter (D): The diameter of the feature around which the image is to be cropped. This helps in determining the range for the cropping offsets.

• Crop Region Size (W): A fixed dimension that defines the size of the square crop region.

+ Output:

• Cropped image region.

• New center coordinates *Cnew*.

Step 1: Initialization
- $W \leftarrow 64$ (crop region size)
- $H \leftarrow W/2$ (half region size for crop)

Step 2: Offset Calculation
- Offset Range $\leftarrow [[D/2]-H, H-\lfloor D/2 \rfloor]$

Step 3: Crop Center Randomization
- $Ox, Oy \sim$ Uniform(Offset Range)
- $Cx', Cy' \leftarrow Cx + Ox, Cy + Oy$

Step 4: Boundary Computation
- TopLeft $\leftarrow (Cx' - H, Cy' - H)$
- BottomRight $\leftarrow (Cx' + H, Cy' + H)$

Step 5: Cropping
- Cropped_Image $\leftarrow$ img[TopLeft:BottomRight]

Step 6: Center Adjustment
- $Cnew \leftarrow (Cx - \text{TopLeft}x, Cy - \text{TopLeft}y)$

Figure 11. Key stages of the Generalized Center-Based Image Cropping algorithm.

The algorithm aims to offer a versatile and precise method for image processing, specifically concentrating on creating transformations and randomization in the vicinity of the lung nodule.

The image cropping method starts by determining the center coordinates of the lung nodule based on annotation data. The Uniform distribution is utilized to create random deviations from the defined center to ensure each cropped image displays a specific variation in the position of the lung nodule.

Utilizing the Uniform distribution is essential in this approach to guarantee unpredictability in the image cropping process. The Uniform distribution is utilized to calculate random variations from the established midpoint, using the formula provided below:

$$P(X = x) = \frac{1}{b-a+1} \qquad (7)$$

X is a random variable with values between a and b, and P(X=x) is the chance of X taking a certain value x within this range. The algorithm applies the deviation, identifies the cropping area, and then crops the image according to the given parameters. The cropping area is predetermined to ensure that the lung nodule stays inside the boundaries of the cropped image. The technique outputs a cropped image and the updated center coordinates. This variability guarantees that each cropped image displays a different location of the lung nodule, which is essential for enhancing diversity in subsequent image analysis and processing (refer to the figure 13).

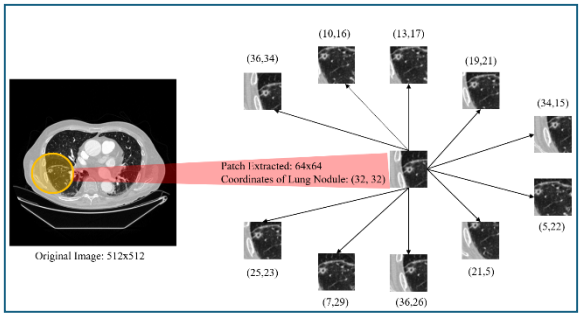

Figure 12. The images generated from the original annotated images.

c. Dataset Division

Following the implementation of the image data enrichment technique, we acquired 11,847 photos measuring 64x64 pixels, showcasing lung nodules ranging in size from 3mm to over 10mm. We have standardized the quantity of lung nodules in our dataset. The data has been categorized into three categories according to the size of the lung nodules. Category 1 contains lung nodules smaller than 5 mm, Category 2 consists of lung nodules ranging from 5 to less than 10 mm, and Category 3 pertains to lung nodules measuring 10 mm or more (refer to the Table 1 and figure 14).

Table 1. The data is divided into three lung nodule size groups.

Categories	Diameter (d) (mm)	Nodule Count	Train	Test
1	d<5	2700	2330	370
2	5<=d<10	6330	5320	1010
3	d>=10	2817	2197	620
	Total	11,847	9847	2000

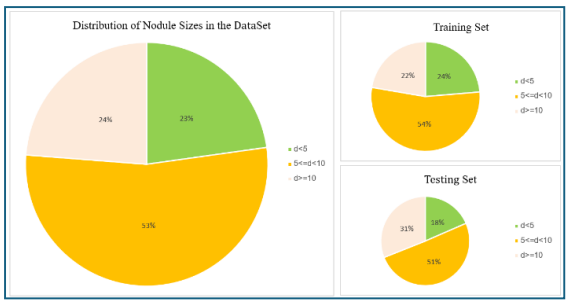

Figure 13. Nodule Count by Diameter Chart.

4.2. Experimental Environment

• We experimented with four deep learning models in the Google Colab Pro environment, a robust and versatile cloud computing platform. The environment has advanced hardware resources, such as around 25 GB of RAM and an Nvidia P100 GPU.

• Optimizer Adam: Adam is a highly effective optimization technique commonly employed in deep learning models. It merges the benefits of Momentum and RMSprop optimization approaches. Adam adapts the learning rate by utilizing real-time calculations of the gradient's first moment (mean) and second moment (unbiased variance). This enables the model to adjust well to the characteristics of the input data and the configuration of the error space.

• Learning Rate Scheduler: decreases the learning rate by a factor of 0.1 after every 30 epochs. Reducing the learning rate enables the model to refine its parameters more effectively in ideal regions and prevents it from missing optimal spots caused by a learning rate that is too high.

4.3. Loss Function

4.3.1. Binary Cross Entropy Loss (BCE Loss)

BCE Loss represented by the formula:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i.\log(p_i) + (1 - y_i).\log(1 - p_i)] \quad (8)$$

where N is the number of samples, this loss function measures the difference between the anticipated value (*pi*) and the actual label (*yi*) for each pixel, ensuring sensitivity for pixel-level distinction.

The gradient of BCELoss with respect to the prediction *pi*

$$\frac{\partial L_{BCE}}{\partial p_i} = -\frac{y_i}{p_i} + \frac{1-y_i}{1-p_i} \quad (9)$$

4.3.2. Dice Loss

Dice Coefficient, a part of Dice Loss, is calculated using the formula:

$$DiceCoef = \frac{2.\sum_{i=1}^{N} p_i y_i + \varepsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + \varepsilon} \quad (10)$$

$$L_{Dice} = 1 - DiceCoef \quad (11)$$

The gradient of Dice Loss with respect to the prediction *pi*

$$\frac{\partial L_{Dice}}{\partial p_i} = \frac{2.(y_i \sum_{i=1}^{N} p - p_i \sum_{i=1}^{N} y) - 2p_i(\sum_{i=1}^{N} y_i p_i + \varepsilon)}{(\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + \varepsilon)^2} \quad (12)$$

where ε is a small constant used to ensure numerical stability. The Dice Coefficient quantifies the similarity between the predicted and actual segmentations by focusing on optimizing the overlapping region between them.

4.3.3. BCEDice Loss

The BCEDice Loss is a composite loss function that aims to utilize the advantages of Binary Cross-Entropy Loss and Dice Loss in training medical image segmentation models.

$$BCEDiceLoss = BCELoss + L_{Dice} \quad (13)$$

The gradient of the combined loss function with respect to the prediction will be the sum of the gradients from both components.

$$\frac{\partial L_{BCEDice}}{\partial p_i} = \frac{\partial L_{BCE}}{\partial p_i} + \frac{\partial L_{Dice}}{\partial p_i} \quad (14)$$

The focus is on reducing loss via BCELoss and improving segment overlap using the Dice Coefficient. This method allows the combined loss function to enhance prediction accuracy at the pixel level while also considering the object's general structure for segmentation.

4.3.4. Metric

The DICE index, also known as the Dice Similarity Coefficient, is utilized to assess the effectiveness of image segmentation algorithms, particularly in the context of medical imagine segmentation.

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (15)$$

This metric measures the overlap of pixels between the segmentation results generated by the model (A) and the ground truth (B). A high DICE index value signifies a strong resemblance between the two segmenters, which is crucial for creating a dependable medical picture segmentation model.

4.4. Multi-Stage Training Techniques for CT Nodule Segmentation

We utilized the Multi-Phase Training technique on lung nodule CT images during the research study. This approach aims to enhance the deep learning model's recognition and analysis capabilities by undergoing two training stages with distinct characteristics and objectives.

Training commences with CT images of 512x512 pixels in size. This step aims to enable the model to acquire comprehensive knowledge of the CT image, encompassing lung structure, surrounding context, and many types of abnormal regions. Utilizing ResUnet++'s pretrain parameter set reduces the time required for the model to adjust to the data and promptly attain fundamental discrimination capability.

The second stage involved training the model on 64x64 pixels CT images that were diced and contained lung nodules. During this phase, the model is improved using checkpoints from the initial phase to better recognize and analyze specific properties of lung nodules. Transitioning from a big to a tiny picture size enables the model to concentrate on acquiring intricate and precise characteristics, hence improving the model's accuracy and classification capability when examining lung nodule images.

4.4.1 Training Stage 1: Feature Exploration in Comprehensive Context

During the first stage of the Multi-Phase Training process, different versions of the PCAM-ResUnet model, such as PAM, CAM, PCAM, and the original ResUnet++ model, are trained using the *practical 512*x*512 dataset*. The objective is to assess the capacity of each model to collect fundamental characteristics from the data (refer to the figure 16). The process commences with the pretraining parameter set of the ResUnet++ model.

The ResUnet++ model demonstrates a consistent and rapid reduction in loss, starting at 0.6208 and steadily falling to 0.0437 by epoch 100. This demonstrates the capacity to promptly seize. The fundamental features of this paradigm are enhanced by the benefits of pretraining. PCAM-ResUnet variants, comprising PCAM, CAM, and PAM, with initial losses of 0.7532, 0.8795, and 0.7208, respectively, consistently show progress with each epoch. While not as fast

as ResUnet++ in reducing loss, all three models are progressively adjusting and effectively learning from the data's properties, reaching losses of 0.1759, 0.1763, and 0.193 correspondingly by epoch 100.

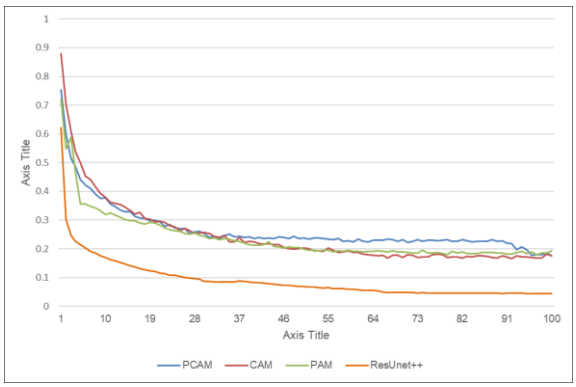

Figure 14. Training Loss for Practical 512x512 Dataset.

4.4.2. Training Stage 2: Focused Refinement on Nodule-Centric Zones

After training on large-size images in the initial phase, we progressed to stage 2 to fine-tune the models using the *Practical 64x64 dataset*. The purpose of this stage is to enhance the capacity to identify intricate characteristics of lung nodules, crucial for medical diagnosis (refer to the figure 16, 17).

The ResUnet++ model began phase 2 with a loss of 0.6208, which rapidly decreased to 0.2147 after the initial 5 epochs. The model demonstrated a notable decrease in loss, reaching 0.0437 by the end of the phase, showcasing its advanced deep learning capacity and efficient optimization. During stage 2, the PCAM, CAM, and PAM models showed notable convergence and optimization abilities. PCAM began with a loss of 0.5508, decreased to 0.215 after 6 epochs, and finally settled at 0.0504, demonstrating consistent convergence. The CAM model initially had a loss of 0.5822, decreased to 0.2052 after 6 epochs, and further improved, demonstrating its adaptability and ongoing learning, achieving a final loss of 0.0552. Unlike the other models, PAM began with a low loss of 0.3632 but had a slower convergence rate, decreasing to 0.2415 after the first 5 epochs and eventually reaching 0.0798, falling short of its initial promising performance. Each model demonstrated its proficiency in deep learning and optimization to different extents, as seen by the decrease in loss throughout training.

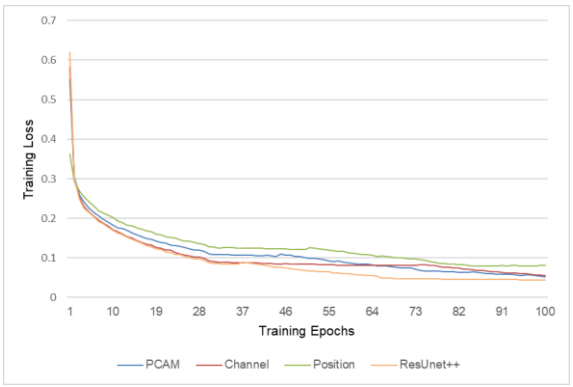

Figure 15. Training Loss for Practical 64x64 Dataset.

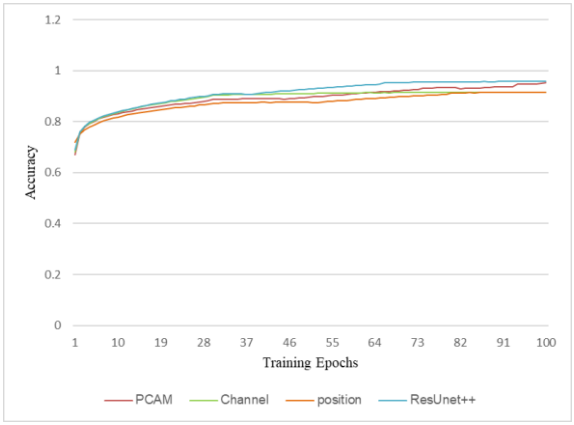

Figure 16. Accuracy for Practical 64x64 Dataset.

4.5. Experimental results

4.5.1 Overall outcomes

The table 2 provides a first look at a comprehensive comparison of the performance of four distinct deep learning models in medical image segmentation tasks, as evaluated by their respective loss values and the Dice Similarity Coefficient (DSC). This overview serves as an initial guide to the nuanced capabilities of each model within the scope of the study.

The ResUnet++ model has superior performance with a little loss of 0.044, indicating effective learning throughout the training process. The model's average Dice Similarity Coefficient (DSC) is 83.89%, which is not the highest. PCAM, while not having the smallest loss, obtains the highest mean DSC at 85.96%. Despite a slightly higher loss value of 0.0529, PCAM is quite effective in capturing the intricacies of medical image segmentation. Its highest observed DSC at 99.64% also underscores its robust performance in the best-case scenarios.

The results table 2 shows a close competition amongst the models in terms of complexity, as shown by the parameter count. Despite incorporating multiple attention modules, PCAM experiences just a slight rise in the number of parameters compared to the individual attention modules when used separately. The use of attention mechanisms in a model can significantly impact performance more than the total number of parameters. The ResUnet++ model has superior performance with a little loss of 0.044, indicating effective learning throughout the training process. The model's average Dice Similarity Coefficient (DSC) is 83.89%, which is not the highest. PCAM, while not having the smallest loss, obtains the highest mean DSC at 85.96%.

The study uses a new method to classify medical images into outcomes according to the Dice Similarity Coefficient (DSC) into four assessment levels: Excellent, Good, Acceptable, and Needs Improvement. The levels correspond to DSC thresholds of equal to or greater than 80%, 60-79%, 40-59%, and less than 40%, respectively. The "Excellent" score signifies a strong correlation between the segmentation findings and the reference standard. The "Good" rating indicates a high level of accuracy that is usually adequate for clinical choices, but there is still potential for enhancement. "Acceptable" indicates that the segmentation outcomes are not perfect but still practical in some situations. Results labeled as "Needs Improvement" with a DSC below 40% necessitate further evaluation and improvement. This method aids in evaluating the segmentation results against expectations and establishes a structure for comparing and enhancing the performance of segmentation algorithms (refer to the figure 18, 19).

The table 3 shows that the PCAM model showed excellent performance in capturing detailed segmentation, with 82.75% of its results achieving the 'Excellent' level out of 2000 test samples evaluated. The ResUnet++ model achieved a high percentage of 'Excellent' outcomes (78.35%) but had a slightly larger proportion of results in the 'Needs Improvement' category, indicating some cases of lower-quality segmentations. The CAM model produced a significant amount of 'Excellent' segmentations (80.35%) but also had a greater proportion of 'Needs Improvement' cases (3.80%) than PCAM, indicating possible inconsistencies in segmentation quality. On the other hand, the PAM model received a significant number of 'Excellent' ratings (77.50%) but also had the greatest rate of 'Needs Improvement' results at 5.15%, suggesting a higher level of variability and more instances of inferior segmentation quality (refer to the figure 20).

Table 2. Comparative Study of Deep Learning Models for Medical Image Segmentation

Model	Loss	Avg DSC	Highest DSC Observed	Parameters
PCAM	0.0529	85.96%	99.64%	14,700,802
CAM	0.0552	84.44%	99.14%	14,479,879
PAM	0.0804	82.79%	99.41%	14,507,039
ResUnet++	0.044	83.89%	99.43%	14,482,564

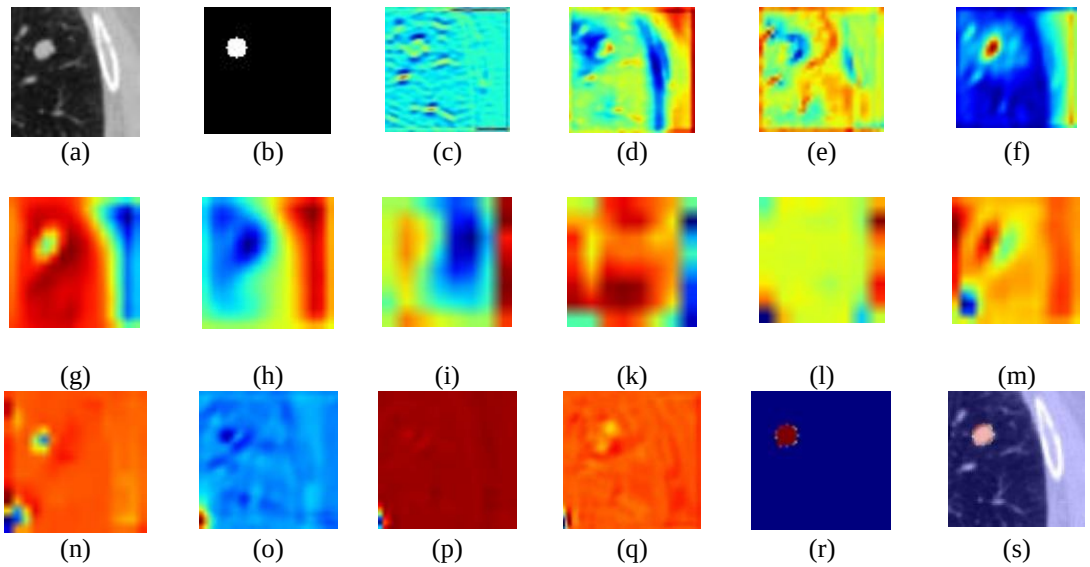

Figure 17. Model Segmentation Visualization Stages with Heatmap Analysis
(a) Input CT image; (b) Ground truth mask; (c-f) Feature maps from early to deep attention layers.
(g-h) Residual features from intermediate layers; (i-k) Outputs from dual attention layers.
(l) ASPP bridge features; (m-o) Decoder outputs at different levels; (p) Enhanced features.
(q) Final segmentation result; (r) Mask on ground truth; (s) Final overlay.

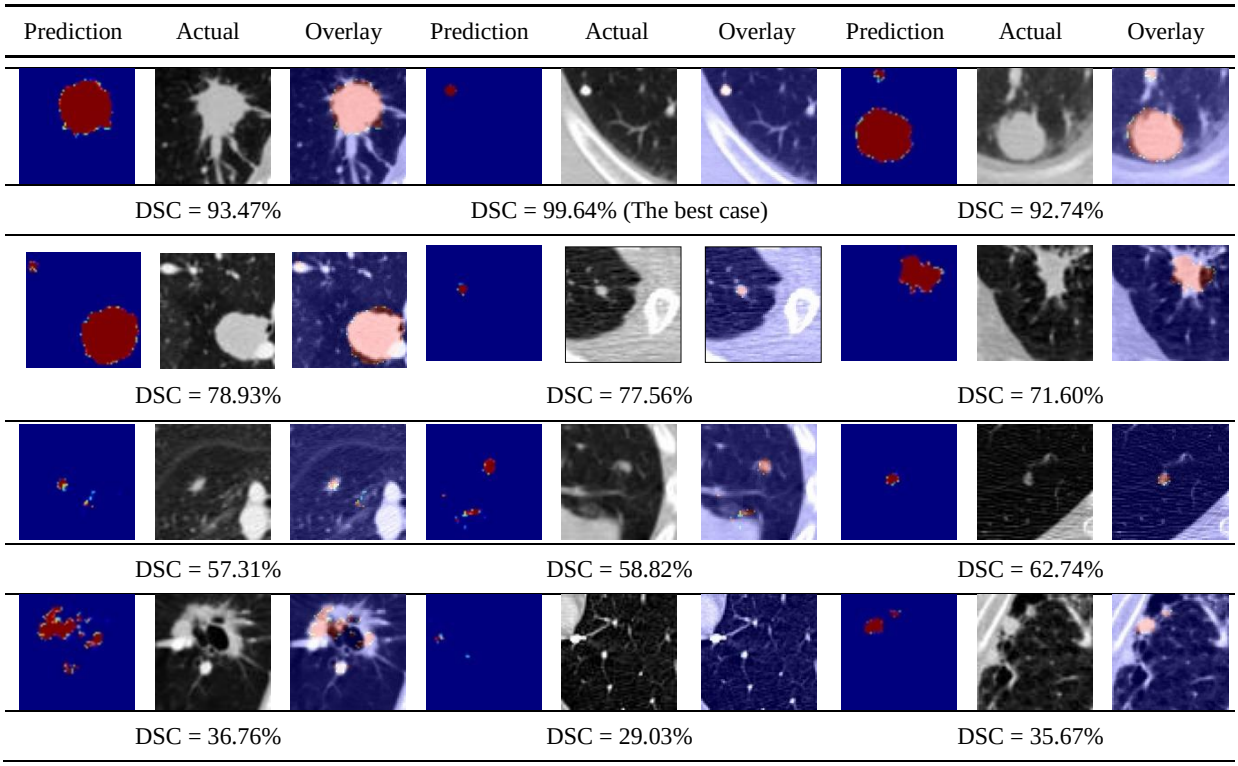

Figure 18. Sample Cases of Segmentation Outcomes Categorized by Achieved DSC Levels.

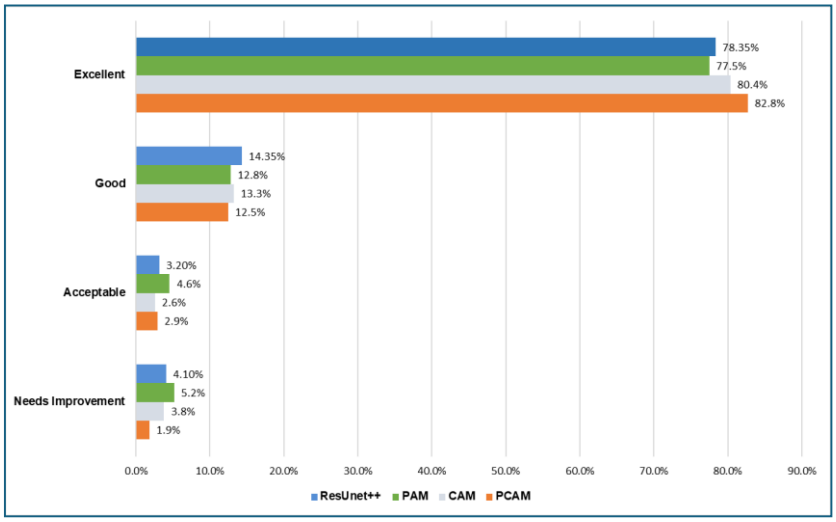

Figure 19. Distribution of DSC Levels Across Deep Learning Segmentation Models.

Table 3. Performance Evaluation of Deep Learning Models on Segmentation Tasks based on DSC Levels

Level	DSC Range	Sample Count			
		Model			
		PCAM	CAM	PAM	Res-Unet++
Needs Improvement	0%<= DSC <40%	37	76	103	82
Acceptable	40%<= DSC <60%	58	52	91	64
Good	60%<= DSC <80%	250	265	256	287
Excellent	80%<= DSC <=100%	1655	1607	1550	1567
Total		2000	2000	2000	2000

4.5.2. Segmentation Accuracy Analysis by Nodule Size Categories

PCAM outperforms in segmentation accuracy across all nodule size categories, with mDSC scores of 82.82% for nodules smaller than 5 mm, 86.64% for nodules between 5 and 10 mm, and 86.71% for nodules 10 mm or bigger in the summarized performance comparison (refer to the Table 4). CAM and ResUnet++ both achieve high performance in segmenting bigger nodules, with CAM scoring 85.34% and ResUnet++ scoring 85.67%. PAM demonstrates enhanced performance as nodules grow in size, achieving its highest mDSC of 84.60% in the largest nodule group. The results emphasize PCAM's outstanding effectiveness in accurately segmenting lung nodules of different sizes.

a. Segmentation Efficacy for Sub-5mm Pulmonary Nodules

The evaluation of 370 samples, including nodules smaller than 5 mm, showed that the PCAM variant of PCAM-ResUnet received a 'Excellent' rating for 71.62% of the samples (refer to the figure 21). This showcases the model's remarkable ability to accurately segment small lung nodules. The sample improvement rate is 2.70%, which is lower than the CAM rate of 4.32% and the PAM rate of 8.92%. The

ResUnet++ model obtained a 64.59% 'Excellent' rate and showed an undesirable gain in the 'Acceptable' rate, with 6.76% of samples falling into this category, surpassing PCAM's 1.89%. This suggests limitations in the original model's ability to deal with smaller and more complex lung nodules.

b. Segmentation Efficacy for 5-10mm Pulmonary Nodules

PCAM outperformed the other models in segmenting pulmonary nodules sized between 5 to 10 mm, achieving 84.26% of instances graded as 'Excellent' (refer to the figure 22) by testing 1010 samples. This variation not only exceeds the others but also excels the original ResUnet++ model significantly, which achieves a rate of 77.43% for samples at the same level.

When examining the 'Needs Improvement' category, PCAM has the lowest rate at 1.88%, while CAM has 2.67% and ResUnet++ has 3.86%. PAM has the highest rate in this category, standing at 4.16%. Models in the 'Acceptable' ($40\% \leq DSC < 60\%$) and 'Good' ($60\% \leq DSC < 80\%$) categories do not exhibit notable distinctions.

c. Segmentation Efficacy for ≥10mm Pulmonary Nodules

Evaluating 620 test samples of lung nodules measuring 10 mm and larger (refer to the figure 23), PCAM demonstrates good accuracy with 86.94% of segmentations rated as 'Excellent'. CAM and ResUnet++ have a higher proportion of 'Excellent' segments in comparison to PCAM. The total percentage of 'Good' and 'Excellent' segments is 92.26% for CAM and 91.94% for ResUnet++. The figures do not surpass the combined total of 93.06% for PCAM. The 'Needs Improvement' rates for CAM and ResUnet++ are 5.32% and 5.00%, respectively, both higher than PCAM's rate of 1.29%.

4.5.3. Comparative Analysis of Methods and Results

Table 5 juxtaposes the efficacy of several lung nodule segmentation models from 2018 to the present, illustrating progress in the incorporation of deep learning architectures and attention mechanisms in medical segmentation endeavors. Conventional U-Net-based models, shown as U-Net (2018) by Tong et al.[18], attained an average Dice Similarity Coefficient (DSC) of 82.05%, underscoring the constraints of U-Net in the absence of attention processes. Subsequent research, including U-Det (2020) by Keetha et al. [13] and the Bidirectional Feature Network (2023) by Sekhara et al. [40], enhanced performance, achieving a DSC of 82.82% by the utilization of advancements such as Bi-FPN and bidirectional topologies.

Integrating attention modules has resulted in substantial enhancements. For example, the Dual-branch Network (2021) by Wu et al [36]. attained an average DSC of 83.16%, but the Dual Encoding Fusion Network (2022) by Xu et al [38]. obtained 85.27%, illustrating that the integration of attention improves the capacity to concentrate on essential characteristics in CT images. Additionally, DA-Net (2021) by Maqsood et al. [39] employed a dual-attention mechanism to enhance both spatial and channel-wise feature representation. The model achieved a DSC of 84.00%, reflecting its effectiveness in segmentation tasks. In this investigation, our PCAM-ResUnet model attained an average Dice Similarity Coefficient of 85.96%, surpassing prior methodologies. This outcome distinctly demonstrates the efficacy of including Channel Attention (CAM) and Position Attention (PAM) into ResUnet blocks, allowing the model to fully use positional and channel information.

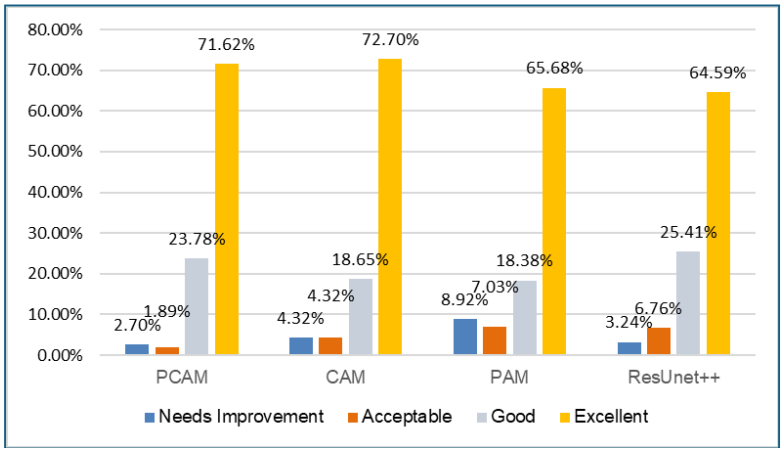

Figure 20. Segmentation Performance Breakdown for Nodules Under 5 mm in Diameter.

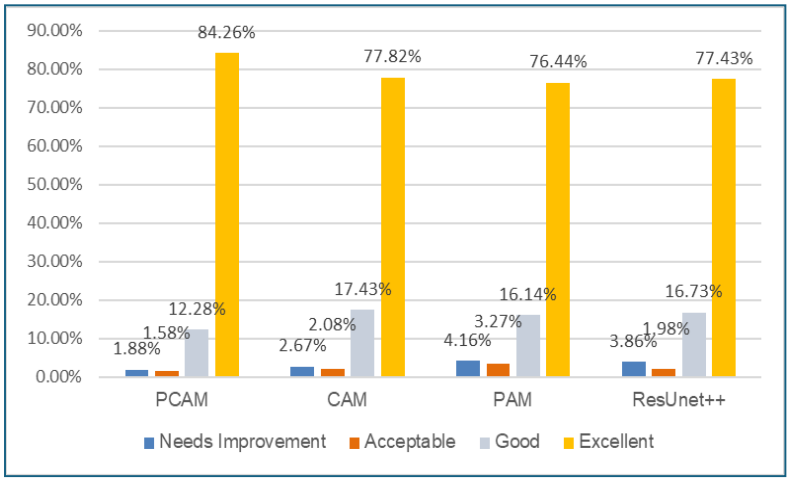

Figure 21. Segmentation Accuracy Analysis for Nodules with Diameter Between 5 and 10 mm.

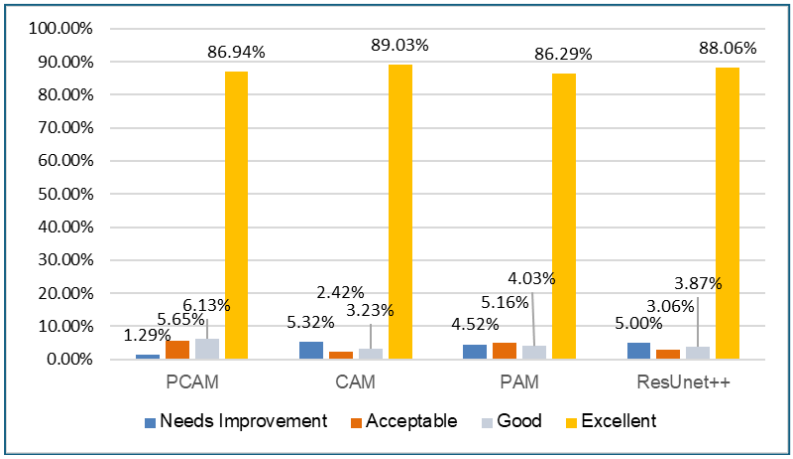

Figure 22. Segmentation Performance Breakdown for Nodules Over 10 mm in Diameter.

Table 4. mDSC Scores for Pulmonary Nodule Segmentation Across Size Categories

Categories	Diameter (d) (mm)	mDSC of Models			
		PCAM	CAM	PAM	ResUnet++
1	d<5	**82.82%**	81.75%	77.95%	80.89%
2	5<=d<10	**86.64%**	84.58%	83.44%	84.40%
3	d>=10	**86.71%**	85.34%	84.60%	85.67%

Table 5. Comparision of segmentation performance

Authors	Model	Dataset	DSC
Tong et al., (2018) [18]	Unet	LIDC	82.05%
Keetha et al., (2020) [13]	U-Det	LUNA16	82.82%
Wu et al., (2021) [36]	Dual-branch network	LIDC	83.16%
Chen et al., (2021) [37]	Fast Multi-crop Guided Attention network	LIDC	81.32%
Xu et al., (2022) [38]	Dual Encoding Fusion Network	LIDC	85.27%
Maqsood et al., (2021) [39]	DA-Net	LIDC	81.00%
Sekhara et al., (2023) [40]	Bidirectional feature network	LUNA-16	82.82%
Our model	**PCAM-ResUnet**	LUNA-16	**85.96%**

5. Discussion

5.1. From SEBlocks to PCAM: Enhancements in PCAM-ResUNet and Their Impact

This discussion will compare and evaluate two primary attention mechanisms: the Squeeze-Excite Block (SEB) and PCAM, which combines the Channel Attention Module (CAM) and Position Attention Module (PAM). The investigation aims to explore how these two strategies improve the medical image segmentation abilities, specifically in recognizing pulmonary nodules in CT scans.

+ Mechanisms and Features of SEB and PCAM

• SEB concentrates on modifying the weights of individual channels by compressing and subsequently enlarging the spatial information. This is achieved by utilizing a pooling layer to compress the spatial information and then passing it through a sequence of fully connected layers to generate attention weights for each channel.

• PCAM integrates channel attention with position attention. CAM uses matrix multiplication and softmax to examine and improve interactions between feature channels, while Position Attention focuses on modeling spatial correlations between pixels. This combination offers a thorough perspective on the image information structure by highlighting crucial feature channels and examining pixel relationships in space.

+ Applications in Medical Image Segmentation

• SEB can enhance accuracy in recognizing crucial features but may not be enough in precisely determining the location and form of critical objects in medical images.

• PCAM has the capability to accurately identify pulmonary nodules in CT images by integrating information from both channel and position. This combination improves the capacity to identify certain characteristics of the nodules and also aids in detecting their location and spatial interactions with nearby structures.

5.2. α and β coefficients

Studying and analyzing the α and β coefficients in the PCAM of our network design show a sophisticated and structured learning mechanism. The coefficients of three PCAM modules inside PCAM-ResUnet exhibit a consistent pattern (refer to the Table 6): PCAM1 has a negative α and positive β, PCAM2 has

both coefficients positive, and PCAM3 has both coefficients negative. This illustrates how the model learns to identify and enhance information from data, as well as how it modifies its focus at various stages.

+ The negative value of alpha and the positive value of beta in PCAM1 indicate an initial spatial information filtering approach, likely aimed at reducing noise and discarding less crucial spatial information while preserving and highlighting key channel information. This could assist in establishing a more precise feature recognition in following layers.

+ PCAM2, with positive coefficients for both space and channel, suggests that integrating input from both aspects is essential during the intermediate phase of the model to capture intricate features and their relationships. This improvement offers a plethora of valuable data to ensure precise segmentation in upcoming stages.

+ Using PCAM3, employing both negative coefficients can serve as a final refining technique to eliminate unnecessary data, emphasize crucial characteristics, and set the stage for the ultimate output reconstruction. This demonstrates the model's self-adjusting method, which involves learning to distinguish useful traits while disregarding irrelevant ones.

We observe that our model learns by progressively amassing information and adjusting its attention to optimize each stage of the learning process. This introduces new avenues for research on how to handle information synthesis in segmentation models.

Table 6. α and βcoefficients for PCAM-ResUnet at different checkpoints

Checkpoint i^{th}	Coefficient	Module		
		PCAM 1	PCAM 2	PCAM 3
i = 98	α	-0.1133	0.0547	-0.0575
	β	0.1705	0.1724	-0.1745
i = 99	α	-0.1116	0.0546	-0.0568
	β	0.1689	0.1686	-0.1761
i = 100	α	-0.113	0.0615	-0.0584
	β	0.171	0.1685	-0.1769

6. Conclusions

A pulmonary nodule with a greater diameter has an increased probability of progressing into a tumor, potentially cancerous. The paper presents PCAM-ResUnet, an enhanced iteration of ResUnet++, designed to increase the accuracy of lung nodule segmentation on CT scans, particularly for nodules measuring 5mm and larger. This model demonstrates the relevance of including various attention modules into the U-shaped design, highlighting the significance of attentive mechanisms in medical image processing.

PCAM-ResUnet performed exceptionally well during the testing stage, achieving an average Dice Similarity Coefficient (DSC) of 85.96% and reaching a peak of 99.64% on the LUNA16 dataset. The results confirm the model's effectiveness in detecting and categorizing potentially tumors and represent a significant advancement beyond current cutting-edge methods.

Nonetheless, the enhancement compared to alternative attention-based techniques is modest, indicating potential for more tuning. Future research will concentrate on investigating attention processes that might dynamically improve the model's capacity to collect essential spatial and contextual attributes. Furthermore,

evaluating the model on datasets with varied features and clinical contexts would facilitate the assessment of its generalizability and adaptability. Expanding the experimental scope aims to demonstrate the model's superiority over existing approaches, optimize its application for medical purposes, and enhance pulmonary nodule diagnosis. Additionally, this work strives to advance deep learning research in the medical field, bridging theoretical advancements with practical clinical applications.

Acknowledgments

This research is funded by University of Science, VNU-HCM under grant number CNTT 2022-20.

References

[1] Siegel, R.L., Miller, K.D., Jemal, A., Cancer statistics, 2020. CA: A Cancer Journal for Clinicians, 70(1), 2020, pp.7–30. doi:https://doi.org/10.3322/caac.21590.

[2] Au-Yong, I.T.H., Hamilton, W., Rawlinson, J. and Baldwin, D.R. . Pulmonary Nodules. BMJ, p.m3673, 2020, doi:https://doi.org/10.1136/bmj.m3673.

[3] B. John Jaidhan and Banavathu Sridevi, A Modified UNet Based Semantic Segmentation Architecture for Pancreas Tumor Detection. International Journal of Bioinformatics Research and Applications, 2024, 20(1). doi:https://doi.org/10.1504/ijbra.2024.10061615.

[4] Cutillo, A., Ganesan, K., Ailion, D.C., Morris, A.H., Durney, C.H., Symko, S.C. and Christman, R, Alveolar Air-tissue Interface and Nuclear Magnetic resonance behavior of lung. Journal of Applied Physiology, 70(5), 1991, pp.2145–2154. doi:https://doi.org/10.1152/jappl.1991.70.5.2145.

[5] Chalela, J.A., Kidwell, C.S., Nentwich, L.M., Luby, M., Butman, J.A., Demchuk, A.M., Hill, M.D., Patronas, N., Latour, L. and Warach, S., Magnetic Resonance Imaging and Computed Tomography in Emergency Assessment of Patients with Suspected Acute Stroke: a Prospective Comparison, The Lancet, [online] 369(9558), 2007, pp.293–298. doi:https://doi.org/10.1016/s0140-6736(07)60151-2.

[6] Gu, Y., Chi, J., Liu, J., Yang, L., Zhang, B., Yu, D., Zhao, Y. and Lu, X, A Survey of Computer-aided Diagnosis of Lung Nodules from CT Scans Using Deep Learning. Computers in Biology and Medicine, 137, 2021, p.104806. doi:https://doi.org/10.1016/j.compbiomed.2021.104806..

[7] Li, R., Xiao, C., Huang, Y., Hassan, H. and Huang, B., Deep Learning Applications in Computed Tomography Images for Pulmonary Nodule Detection and Diagnosis: A Review, Diagnostics, 12(2), 2022, p.298. doi:https://doi.org/10.3390/diagnostics12020298.

[8] Nithila, E.E. and Kumar, Segmentation of Lung Nodule in CT Data using Active Contour Model and Fuzzy C-mean Clustering, Alexandria Engineering Journal, [online] 55(3), 2016, pp.2583–2588. doi:https://doi.org/10.1016/j.aej.2016.06.002.

[9] Mansoor, A., Bagci, U., Foster, B., Xu, Z., Papadakis, G.Z., Folio, L.R., Udupa, J.K. and Mollura, Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends. RadioGraphics, 35(4), 2015, pp.1056–1076. doi:https://doi.org/10.1148/rg.2015140232.

[10] Zhang, J., Xia, Y., Cui, H. and Zhang, Y, Pulmonary Nodule Detection in Medical Images: A survey, Biomedical Signal Processing and Control, 43, 2018, pp.138–147. doi:https://doi.org/10.1016/j.bspc.2018.01.011.

[11] Riquelme, D. and Akhloufi, M.A, Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans. AI, [online] 1(1), 2020, pp.28–67. doi:https://doi.org/10.3390/ai1010003.

[12] Wu, B., Zhou, Z., Wang, J. and Yu, Y, Joint Learning for Pulmonary Nodule Segmentation, Attributes and Malignancy prediction, 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 2018, pp. 1109-1113, doi: 10.1109/ISBI.2018.8363765.

[13] Nikhil Varma Keetha, Samson and Sekhara, C, U-Det: A Modified U-Net Architecture with Bidirectional Feature Network for Lung Nodule Segmentation, arXiv (Cornell University),2020, doi:https://doi.org/10.48550/arxiv.2003.09293.

[14] Zhao, C., Han, J., Yang, J. and Gou, F, Lung Nodule Detection via 3D U-Net and Contextual Convolutional Neural Network, 2018 International Conference on Networking and Network Applications (NaNA), Xi'an, China, 2018, pp. 356-361, doi: 10.1109/NANA.2018.8648753.

[15] Long, J., Shelhamer, E. and Darrell, T, Fully Convolutional Networks for Semantic Segmentation,

2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [online], doi:https://doi.org/10.1109/cvpr.2015.7298965.

[16] Ronneberger, O., Fischer, P. and Brox, T, U-Net: Convolutional Networks for Biomedical Image Segmentation, Lecture Notes in Computer Science, 9351, 2015, pp.234–241.

[17] Huang, X., Sun, W., Tseng, T.-L. (Bill), Li, C. and Qian, W, Fast and Fully-Automated Detection and Segmentation of Pulmonary Nodules in Thoracic CT Scans using Deep Convolutional Neural Networks, Computerized Medical Imaging and Graphics, 74, 2019, pp.25–36. doi:https://doi.org/10.1016/j.compmedimag.2019.02.003.

[18] Tong, G., Li, Y., Chen, H., Zhang, Q. and Jiang, H., Improved U-NET Network for Pulmonary Nodules Segmentation'. Optik, [online] 174, 2018, pp.460–469. doi:https://doi.org/10.1016/j.ijleo.2018.08.086.

[19] Rensink, R.A, The Dynamic Representation of Scenes, Visual Cognition, 7(1-3), 2000, pp.17–42. doi:https://doi.org/10.1080/135062800394667.

[20] Hu, J., Shen, L. and Sun, G, Squeeze-and-Excitation Networks, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp.7132–7141. doi:https://doi.org/10.1109/cvpr.2018.00745.

[21] Qin, Z., Zhang, P., Wu, F. and Li, X., FcaNet: Frequency Channel Attention Networks, IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 763-772, doi: 10.1109/ICCV48922.2021.00082.

[22] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. , ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:https://doi.org/10.1109/cvpr42600.2020.01155.

[23] Lee, H., Kim, H.-E. and Nam, H., 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1854-1862, doi: 10.1109/ICCV.2019.00194.

[24] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., Attention u-net: Learning Where to Look for the Pancreas, 2018, arXiv preprint arXiv:1804.03999

[25] Hu, J., Shen, L., Albanie, S., Sun, G. and Vedaldi, A., Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks, arXiv (Cornell University), 31, 2018, pp.9401–9411. https://doi.org/10.1109/TWC. 2012.083112.120127.

[26] Woo, S., Park, J., Lee, J.-Y. and Kweon, CBAM: Convolutional Block Attention Module. Computer Vision – ECCV 2018, pp.3–19. doi:https://doi.org/10.1007/978-3-030-01234-2_1.

[27] Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z. and Lu, H. (2019). 'Dual Attention Network for Scene Segmentation'. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:https://doi.org/10.1109/cvpr.2019.00326.

[28] Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S. and Pal, The Importance of Skip Connections in Biomedical Image Segmentation, Deep Learning and Data Labeling for Medical Applications, 2016, pp.179–187.

[29] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, UNet++: A Nested U-Net Architecture for Medical Image Segmentation, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, [online], 2018, pp.3–11.

[30] Zhang, Z., Liu, Q. and Wang, Y., Road Extraction by Deep Residual U-Net. IEEE Geoscience and Remote Sensing Letters, [online] IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, 2018, pp. 749-753, doi: 10.1109/LGRS.2018.2802944.

[31] He, K., Zhang, X., Ren, S. and Sun, J., Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770–778. doi:https://doi.org/10.1109/cvpr.2016.90.

[32] B. John Jaidhan and Banavathu Sridevi, A Modified UNet Based Semantic Segmentation Architecture for Pancreas Tumor Detection, International Journal of bioinformatics research and applications, 20(1), 2024. doi:https://doi.org/10.1504/ijbra.2024.10061615.

[33] Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., Lange, T.D., Halvorsen, P. and D. Johansen, H, ResUNet++: An Advanced Architecture for Medical Image Segmentation. 2019 IEEE International Symposium on Multimedia (ISM). doi:https://doi.org/10.1109/ism46123.2019.0004.

[34] He, K., Zhang, X., Ren, S. and Sun, J., Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, Computer Vision – ECCV 2014, pp.346–361. doi:https://doi.org/10.1007/978-3-319-10578-9_23.

[35] Setio, A.A.A., Traverso, A., de Bel, T., Berens, M.S.N., Bogaard, C. van den, Cerello, P., Chen, H., Dou, Q., Fantacci, M.E., Geurts, B., Gugten, R. van der, Heng, P.A., Jansen, B., de Kaste, M.M.J., Kotov, V., Lin, J.Y.-H., Manders, J.T.M.C., Sóñora-Mengana, A., García-Naranjo, J.C. and

Papavasileiou, E., Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography images: The LUNA16 challenge, Medical Image Analysis, [online] 42, pp.1–13. doi:https://doi.org/10.1016/j.media.2017.06.015.

[36] Wu, Z., Zhou, Q. and Wang, F., Coarse-to-Fine Lung Nodule Segmentation in CT Images With Image Enhancement and Dual-Branch Network, IEEE Access, 9, 2021, pp.7255–7262. doi:https://doi.org/10.1109/access.2021.3049379.

[37] Chen, Q., Xie, W., Zhou, P., Zheng, C. and Wu, D, Multi-Crop Convolutional Neural Networks for Fast Lung Nodule Segmentation, IEEE Transactions on Emerging Topics in Computational Intelligence, 2021, pp.1–11. doi:https://doi.org/10.1109/tetci.2021.3051910.

[38] Xu, W., Xing, Y., Lu, Y., Lin, J. and Zhang, X., Dual Encoding Fusion for Atypical Lung Nodule Segmentation, 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). https://doi.org/10.1109/isbi52829.2022.9761405.

[39] Maqsood, M., Yasmin, S., Mehmood, I., Bukhari, M. and Kim, M., An Efficient DA-Net Architecture for Lung Nodule Segmentation, Mathematics 2021, 9, 1457, https://doi.org/10.3390/math9131457.

[40] Sekhara, C., Samson, Nikhil Varma Keetha, Praveen Kumar Donta and Gurindapalli Rajita, A Bi-FPN-Based Encoder–Decoder Model for Lung Nodule Image Segmentation, Diagnostics, 13(8), 2023, pp.1406–1406. doi:https://doi.org/10.3390/diagnostics13081406.