CAAF-RESUNET: DYNAMIC FUSION OF ATTENTION MECHANISMS FOR CONTEXT-AWARE LUNG CT SCAN SEGMENTATION

THANG QUOC PHAM [1], THAI HOANG LE[2,3,*], KHAI DINH LAI[2,3], DAT QUOC NGO[1], TAN VAN PHAM[1], QUANG HONG HUA[1], KHANG QUANG LE[1], HUYEN LE DUY MAI[4], TUYEN NGUYEN NGOC LAM[4]

[1] University of Medicine and Pharmacy at HCMC, Ho Chi Minh City 700000, Viet Nam
[2] University of Science, Ho Chi Minh City 700000, Viet Nam
[3] Vietnam National University, Ho Chi Minh City 700000, Viet Nam
[4] University Medical Center Ho Chi Minh City, Ho Chi Minh City 700000, Viet Nam
[*] Corresponding author: lhthai@fit.hcmus.edu.vn
E-MAIL: phamquocthang@ump.edu.vn, lhthai@fit.hcmus.edu.vn, laidinhkhai@sgu.edu.vn, ngoquocdat@ump.edu.vn, pvtan.ntgpb21@ump.edu.vn, huahongquang@ump.edu.vn, khanglequang@ump.edu.vn, huyen.ldm@umc.edu.vn, tuyen.nnl@umc.edu.vn

Abstract:

Segmentation of pulmonary nodules in CT scans is critical for early diagnosis and treatment of lung diseases. Traditional models like U-Net often face challenges with small nodule sizes, and noisy data. To address these limitations, we propose CAAF-ResUnet (Context-Aware Adaptive Attention Fu-sion ResUnet), which combines Channel Attention and Position Attention using an Adaptive Attention Fusion (AAF) mechanism, controlled by an Adaptive Attention Controller (AAC). Our model achieves commendable performance on the LUNA16 and LUNA-Noise datasets, with Dice Scores of 89.44% and 88.55%, respectively. These results surpass existing methods, demonstrating the model's ability to prioritize meaningful features and handle noisy inputs effectively. CAAF-ResUnet provides a robust solution for nodule segmentation in challenging scenarios, paving the way for improved diagnostic accuracy in clinical settings. This work emphasizes the importance of adaptive attention mechanisms for medical image segmentation and highlights opportunities for future improvements.

Keywords:

Adaptive Attention Controller, Adaptive Attention Fusion, attention mechanism

1. Introduction

Lung cancer is a predominant cause of global death, frequently diagnosed at ad-vanced stages, which markedly diminishes therapeutic efficacy [1]. Early screening with chest CT can detect atypical lung nodules, enabling prompt diagnosis and better patient outcomes. Accurate segmentation of lung nodules aids risk assessment and therapy planning. However, small size, unclear margins, and complex surrounding structures make segmentation challenging. Traditional methods like U-Net and its variants have significantly advanced medical image segmentation. However, models using multiple attention mechanisms have not fully optimized the assessment of each mechanism's impact based on image context. This limitation is especially evident in segmenting tiny nodules, handling ambiguous margins, or dealing with noisy conditions, leading to reduced precision. To tackle these issues, we offer CAAF-ResUNet (Context-Aware Adaptive Atten-tion Fusion U-Net), a model derived from the Res-UNet++ architecture, particularly engineered for the segmentation of pulmonary nodules in CT images. CAAF-ResUNet prominently features the adaptable incorporation of two primary attention mechanisms, channel Attention and Position Attention, through Adaptive Attention Fusion (AAF), regulated by an Adaptive Attention Controller (AAC). This system optimally balances key features and adapts to varying contextual requirements of input images. Key contributions of this study include: (1) AAF and AAC integration: An adaptive weighting method dynamically balances Channel and Position Attention, improving segmentation efficiency. (2) Improved segmentation performance: CAAF-ResUNet achieves a Dice Score of 89.44% on the LUNA16 dataset, outperforming a model without AAC (87.67%), demonstrating the effectiveness of adaptive weighting. (3) Robustness to noise: The model maintains stable performance on a Gaussian noise-augmented LUNA16 dataset, proving its resilience in real-world noisy conditions.

2. Literature review

2.1. Initial Phase: Models Derived from U-Net

U-Net (Ronneberger et al., 2015) [2] established the basis for medical picture seg-mentation with its symmetric encoder-decoder design and skip connections, facili-tating the acquisition of multi-scale features. Subsequently, U-Net++ (Zhou et al., 2019) [3] augmented the model by include layered skip connections, hence enhanc-ing multi-scale representation. Res-UNet (Zhang et al., 2018) [4] included residual blocks to mitigate gradient vanishing problems during training.

2.2. Evolution Phase: Advancements with Attention Mechanisms

The advent of attention mechanisms marked a significant advancement in medical image segmentation, addressing the limitations of U-Net in capturing complex spatial and channel relationships. Attention U-Net [5] introduced a method to dynamically emphasize essential image regions, significantly improving segmentation, particularly in areas with subtle boundaries or low contrast. ResUNet++ [6] further built on this by incorporating Squeeze-and-Excitation (SE) blocks, refining inter-channel dependencies. By enhancing relevant feature channels and suppressing redundant ones, it proved effective in segmenting small and complex structures, such as lung nodules in CT images. Later advancements included Dual Attention Network (DANet) [7], which combined Channel Attention and Position Attention, improving the ability to capture both inter-channel and spatial interactions, thus enhancing segmentation of intricate anatomical structures. AWEU-Net [8] expanded this approach with adaptive weighting algorithms, dynamically balancing feature contributions across encoder-decoder layers, demonstrating strong performance on noisy datasets and challenging conditions.

2.3. Present Phase: Transformer-Driven Models

The advent of Transformer-based models has established a novel paradigm for med-ical picture segmentation. TransUNet [9] integrates Transformer layers with the U-Net framework, facilitating the acquisition of long-range relationships in data. SwinUNet [10] utilises the Swin Transformer to acquire multi-scale characteristics. Although these models demonstrate superior performance on extensive datasets, they need considerable computing resources and ample training data.

2.4. Research gaps

While attention-based segmentation has progressed, integrating multiple attention types dynamically remains challenging. Most existing methods apply static attention, lacking input adaptability. Transformer-based models, though powerful, are often impractical in clinical settings due to high computational demands. Thus, there is a need for adaptive attention coordination that balances performance and efficiency, especially under limited or noisy data conditions.

3. Method

3.1. Overview of Context-Aware Adaptive Attention Fusion ResUNet (CAAF-ResUnet)

CAAF-ResUNet (Context-Aware Adaptive Attention Fusion ResUNet) is a model explicitly engineered for lung nodule segmentation in CT images, founded on the Res-Unet++ architecture [6]. This model incorporates two attention mechanisms—Channel Attention and Position Attention—alongside an adaptive weighting mecha-nism, facilitating the prioritizing of essential aspects according to the input image's context. The CAAF-UNet design consists of three primary stages: Encoder, Bridge, and Decoder (see Figure 1)

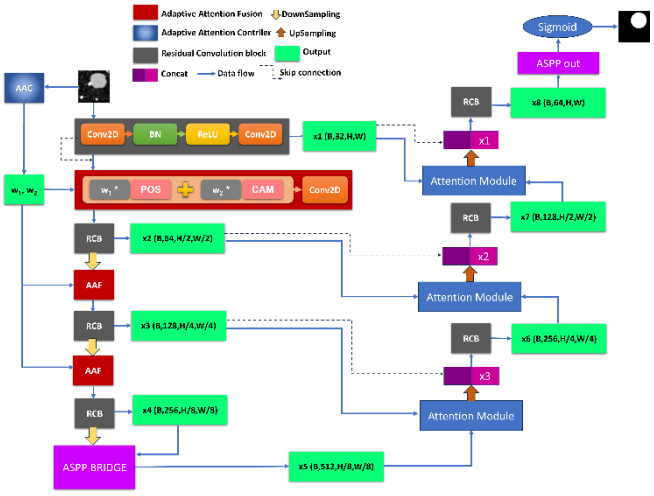


FIGURE 1. Comprehensive architecture of the CAAF-ResUnet model.

- The encoder enhances CT feature extraction using Residual Convolutional Blocks, Batch Normalization, and ReLU for stable training and minimal information loss. Residual connections preserve key features, while the hierarchical structure captures local textures (64×64×32) and global context (8×8×256). A central

component, Adaptive Attention Fusion (AAF), adaptively balances Channel and Position Attention via the Adaptive Attention Controller (AAC), which adjusts weights based on image context.

- The Bridge leverages Atrous Spatial Pyramid Pooling (ASPP) to extract multi-scale features between the encoder and decoder. By applying convolutions with dilation rates of 1, 6, 12, and 18, along with global average pooling, ASPP captures both local details and global context. A 1×1 convolution refines the output to 8×8×256, maintaining spatial dimensions.
- The decoder merges bridge and encoder features at each stage, enhanced by AAF, and progressively upsamples to 64×64×32. A final 1×1 convolution followed by Sigmoid produces the segmentation map with values in [0, 1].

3.2. Channel Attention module

Considering the input $X \in \mathbb{R}^{B\times C\times H\times W}$, Channel Attention initially converts each spatial channel of X into distinct feature vectors F_a^c of size $H\times W$ by flattening the spatial dimensions. The feature vectors are subsequently aggregated along the channel dimension to create a feature matrix F_a.

$$F_a = reshape(X,(B,C,H\text{x}W)) \quad (1)$$

The attention correlation matrix A among channels is calculated as the matrix product of F_a and its transposed variant F_a^T:

$$A = F_a.F_a^T \quad (2)$$

To minimize dominating channels, the attention matrix A is changed by deleting the largest value in each row and normalizing using the softmax function to assure weights in each row equal to 1:

$$A_{norm} = softmax(\max(A,-1) - A,-1) \quad (3)$$

Applying the attention matrix A_{norm} on F_a to generate the focused feature F_e, which is then transformed to its spatial dimensions:

$$F_e = reshape(A_{norm}.F_a,(B,C,H,W)) \quad (4)$$

Ultimately, the concentrated feature Fe is integrated with the original input X via the learnt parameter β, which modulates the impact of the attention mechanism:

$$F_{out} = \beta.F_e + X \quad (5)$$

3.3. Position Attention module

Position Attention is intended to improve the spatial connections among pixels, allowing the model to precisely discern essential features such as borders and intricate areas within the image. The input $X \in \mathbb{R}^{B\times C\times H\times W}$ is downsampled by max pooling with an 8×8 kernel size, emphasising essential characteristics and minimising computational expenses.

X is first transformed by two 1×1 convolutions into spatial feature vectors F_b and F_c, representing two complementary views. Fb is reshaped to [B,H×W,C'], capturing position-wise relationships, while Fc is reshaped to [B,C',H×W], preserving channel-wise details. These representations are fed into a correlation matrix, where each entry (*i, j*) reflects the spatial similarity between pixels *i* and *j*. The resulting matrix, normalized by Softmax, highlights strong spatial dependencies and suppresses weak connections, enabling the model to enhance structure-aware learning and boundary discrimination.

$$A_s = softmax(F_b.F_c) \quad (6)$$

The spatial attention feature Fe is calculated via

$$F_e = Conv(X).A_s^T \quad (7)$$

Finally, Fe is interpolated back to the original dimensions of X and combined with X utilising the tuning parameter α to get the final output:

$$F_{out} = \alpha.F_e + X \quad (8)$$

3.4. The Adaptive Attention Controller (AAC)

The Adaptive Attention Controller in CAAF-ResUNet is essential for harmonizing Channel Attention and Position Attention, allowing the model to dynamically modify weights according on the input image's context. The procedure starts with the condensation of information from the input image $X \in \mathbb{R}^{B\times C\times H\times W}$. Employing Global Average Pooling generates a global feature vector $g \in \mathbb{R}^{B\times C}$, which captures the comprehensive channel information. The vector g is further processed through two fully connected layers to get the weights w_1 and w_2, which signify the priorities of Position Attention and Channel Attention, respectively. This procedure is mathematically expressed as:

$$z = ReLU(W_1 g + b_1), \quad (9)$$

$$w = Softmax(W_2 + b_2) \quad (10)$$

Here, $W1$ and $W2$ denote the weight matrices, whereas $b1$ and $b2$ represent the bias vectors of the fully connected layers. The softmax function normalizes the weights w_1 and w_2, ensuring that $w_1 + w_2 = 1$, which represents the distribution of attention priorities. The weights are subsequently utilized to amalgamate the characteristics from Position Attention (F position) and Channel Attention (F channel) as follows:

$$F_{encoded} = w_1 \times F_{position} + w_2 \times F_{channel} \quad (11)$$

The AAC dynamically adjusts attention weights to suit each image's context, emphasizing spatial cues (w_1) for small nodules or unclear borders, and channel cues (w_2) when spectral features dominate. This adaptability enhances segmentation accuracy and generalization. As a core of CAAF-ResUNet, AAC ensures an effective balance between flexibility and performance.

3.5. Adaptive Attention Fusion (AAF)

Adaptive Attention Fusion (AAF) combines Position and Channel Attention using adaptive weights w_1 and w_2 generated by the Attention Controller (AAC). Given input X ∈ [B, C, H, W], both attentions are applied separately to produce Position Attention (P) and Channel Attention(C). These outputs are then fused using the weights, which are broadcasted to match spatial dimensions, allowing AAF to adaptively emphasize spatial or channel features based on image context.

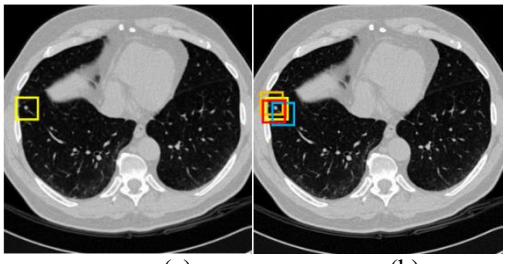
$$F_{fused} = w_1 \times P + w_2 \times \mathrm{C} \quad (12)$$

The fused tensor F_{fused} is refined by a 3×3 convolutional, producing an optimized feature map that combines spatial and channel information. AAF enables effective modulation of attention, helping the model capture both local details and global context.

4. Experimental Environment

4.1. Dataset

This study uses two datasets derived from LUNA16 [11] for pulmonary nodule segmentation. The original LUNA16 includes 1,186 annotated nodules from 888 CT scans. LUNA-Noise is generated by adding Gaussian noise (mean = 0, variance = 0.001) to simulate real-world conditions and assess model robustness. Using nodule center coordinates, 64×64 patches were extracted to ensure full coverage. Both datasets were identically split into 9,800 training and 1,200 testing patches for consistent evaluation across clean and noisy data.



(a) (b)

FIGURE 2. Illustration of Pulmonary Nodule Patch Generation. (a) Nodule center highlighted by the yellow bounding box based on the annotation file. (b) Generated example patches (orange, blue, red bounding boxes)

4.2. Model Implementation

Models were trained in PyTorch on an NVIDIA A100 using AdamW (lr=0.0001, weight decay=0.01). A StepLR scheduler reduced the learning rate by 0.1 every 30 epochs, with a lower bound of 1×10^{-6} to prevent freezing. The loss combined BCE and Dice, with batch size 8 and 100 training epochs.

4.3. Evaluation Metrics

This study evaluates lung nodule segmentation using Dice, IoU, Sensitivity, Specificity, and Miss Rate, all derived from the confusion matrix. TP and TN indicate correct segmentation of nodule and non-nodule pixels, while FP and FN reflect over- and under-segmentation. Together, these metrics provide a comprehensive view of model accuracy and error characteristics.

$$Dice = \frac{2\times TP}{(2\times TP+FN+FP)} \quad (13)$$

$$IoU = \frac{TP}{(TP+FN+FP)} \quad (14)$$

$$Sensitivity = \frac{TP}{(TP+FN)} \quad (15)$$

$$Miss\ Rate = \frac{FN}{(TP+FN)} \quad (16)$$

$$Specificity = \frac{TN}{(TN+FP)} \quad (17)$$

5. Experimental Results

5.1. Ablation Study: Evaluating the Effectiveness of Adaptive Attention Integration

This section outlines an ablation research designed to thoroughly assess the effects of substituting and incorporating different attention methods into the basic ResUnet++ architecture. To systematically evaluate the impact of attention processes, we incrementally adjust the model in Table 1:

TABLE 1. Configuration of ablation study models

Model	Description
Position+ResUnet	Replacing the SEBlock in ResUnet++ with Position Attention
Channel+ResUnet	Replacing the SEBlock with Channel Attention
Dual-ResUnet	Integrating both Channel and Position Attention without using AAC for adaptive weighting.
CAAF-ResUnet	Extending the Dual-ResUnet by introducing the Adaptive Attention Controller (AAC) to dynamically fuse channel and position Attention.

5.2. Performance on LUNA Dataset

Table 2 presents ablation results on the LUNA dataset. The baseline ResUnet++ achieves Dice 85.69% and IoU 75.63%. Replacing SEBlocks with Position or Channel Attention yields similar performance, with Channel + ResUnet showing slightly higher Specificity (99.46%) but both models share a high Miss Rate (~14.55%). Dual-ResUnet, combining both attentions without weighting, improves Dice to 87.67% and reduces Miss Rate to 13.13%. CAAF-ResUnet, using dynamic attention via AAC, achieves the best performance (Dice 89.44%, IoU 81.23%) with a lower Miss Rate of 10.16%. The learned weights ($w_1 = 0.29$, $w_2 = 0.71$) indicate a stronger

reliance on Channel Attention in clean CT data.

TABLE 2. Ablation study model performance on Luna dataset

Model	Dice	IoU	Sensitivity	Specificity	Miss Rate
ResUnet ++(Baseline)	85.69%	75.63%	86.36%	99.35%	14.86%
Position + ResUnet	85.19%	74.75%	85.45%	99.34%	14.55%
Channel + ResUnet	85.41%	75.15%	85.08%	99.46%	14.92%
Dual-ResUnet	87.67%	78.45%	86.87%	99.51%	13.13%
CAAF-ResUnet	**89.44%**	**81.23%**	**89.84%**	**99.56%**	**10.16%**

5.3. Performance on LUNA-Noise Dataset

On the LUNA-Noise dataset, the baseline ResUnet++ shows reduced robustness, with Dice and IoU dropping to 85.30% and 75.03%, respectively. Position + ResUnet and Channel + ResUnet perform similarly but suffer from higher Miss Rates, notably 15.01% for Channel + ResUnet, indicating difficulty under noisy conditions. Dual-ResUnet improves resilience with Dice 87.37%, IoU 77.99%, and a lower Miss Rate of 11.46%, confirming the benefit of combining both attention types. CAAF-ResUnet outperforms all, achieving Dice 88.55%, IoU 79.80%, and the lowest Miss Rate (10.19%), with adaptive weighting favouring position attention (w_1 = 0.37), highlighting its importance in handling spatial noise.

TABLE 3. Ablation study model performance on Luna-Noise dataset

Model	Dice	IoU	Sensitivity	Specificity	Miss Rate
ResUnet ++(Baseline)	85.30%	75.03%	85.13%	99.39%	14.87%
Position + ResUnet	85.26%	74.83%	85.38%	99.32%	14.62%
Channel + ResUnet	85.43%	75.15%	84.99%	99.44%	15.01%
Dual-ResUnet	87.37%	77.99%	86.45%	99.51%	11.46%
CAAF-ResUnet	**88.55%**	**79.80%**	**87.82%**	**99.52%**	**10.19%**

5.4. Comparative Analysis with Existing Models

Table 4 provides a comparative assessment of segmentation efficacy between CAAF-ResUnet and other leading algorithms on the LIDC and LUNA16 datasets. Our suggested CAAF-ResUnet attains a Dice Similarity Coefficient (DSC) of 89.44% on the LUNA16 dataset, exceeding all other models by a considerable margin. In the LUNA16 evaluation, Keetha et al. [13] and Sekhara et al. [16] attained a Dice Similarity Coefficient (DSC) of 82.82%, demonstrating superior performance relative to other benchmarks. Nevertheless, these models are deficient in the adaptive fusion capabilities shown in CAAF-ResUnet, constraining their capacity to manage subtle contextual factors. CAAF-ResUnet's efficacy is attributed to its Adaptive Attention Fusion (AAF) mechanism, which amalgamates Channel and Position Attention with dynamic context-aware weighting through the AAC module. This architecture improves segmentation efficacy, especially under noisy or difficult conditions.

TABLE 4. Comparison of lung nodule segmentation models

Authors	Model	DSC
Tong et al., (2018)[12]	Unet	82.05%
Keetha et al., (2020) [13]	U-Det	82.82%
Wu et al., (2021) [14]	Dual-branch network	83.16%
Maqsood et al., (2021) [15]	DA-Net	81.00%
Sekhara et al., (2023) [16]	Bidirectional feature network	82.82%
Our model	**CAAF-ResUnet**	**89.44%**

The Figure 3 presents many cases demonstrating both accurately segmented nodules and occurrences with slight segmentation errors. Notwithstanding these variations, the models effectively identify the locations of pulmonary nodules.

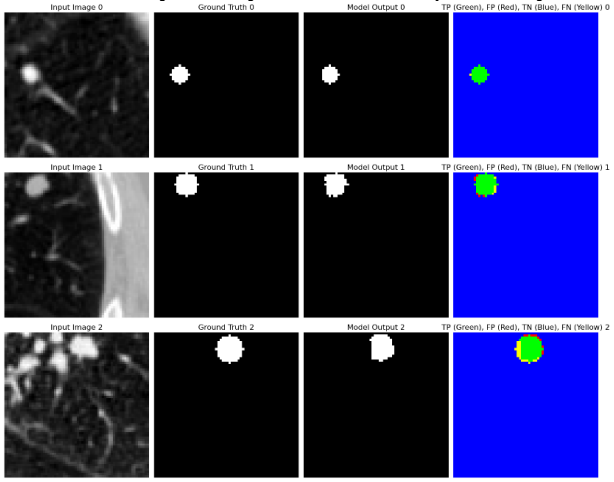


FIGURE 3. Visual results of pulmonary nodule segmentation. From left to right: input image, ground truth, model output, and overlay of TP (green), FP (red), FN (yellow), and TN (blue).

5. Conclusions

This paper presents CAAF-ResUnet, an innovative segmentation model designed for the identification of pulmonary nodules in CT images. Our approach, based on the ResUnet++ framework, incorporates Channel Attention and Position Attention via the Adaptive Attention Fusion (AAF) mechanism, governed by the Adaptive Atten-tion Controller

(AAC). This adaptive weighting method allows the model to priori-tise pertinent elements dynamically, depending on the input image's context, so overcoming the constraints of static attention processes in current models.

Experimental findings on the LUNA16 and LUNA-Noise datasets illustrate the robustness of CAAF-ResUnet, attaining Dice Scores of 89.44% and 88.55%, respectively, and markedly surpassing baseline models and various methodologies. Fur-thermore, the model's capacity to equilibrate false positives and false negatives, as evidenced by its exceptional IoU, Sensitivity, Specificity, and Miss Rate, under-scores its efficacy in managing difficult situations.

Future research will concentrate on the extension of CAAF-ResUnet to other medical imaging tasks, the optimisation of its computational efficiency for deploy-ment in resource-constrained environments, and the investigation of hybrid attention mechanisms to achieve additional performance improvements. The objective of this research is to provide a scalable and robust solution for pulmonary nodule segmenta-tion, thereby enabling early diagnosis and enhanced treatment outcomes in clinical practice.

Acknowledgements

This work has received support from the Korea International Cooperation Agency (KOICA) under the project entitled “Education and Research Capacity Building Project at University of Medicine and Pharmacy at Ho Chi Minh City,” conducted from 2023 to 2025 (Project No. 2021-00020-3).

References

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, “Cancer statistics, 2023,” CA: A Cancer Journal for Clinicians, vol. 73, no. 1, pp. 17–48, 2023.

[2] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” Lecture Notes in Computer Science, vol. 9351, pp. 234–241, 2015.

[3] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11, 2018. [Online]. Available: https://arxiv.org/abs/1807.10165

[4] Z. Zhang, Q. Liu, and Y. Wang, “Road extraction by deep residual U-Net,” IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 749–753, 2018.

[5] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, and B. Kainz, “Attention U-Net: Learning where to look for the pancreas,” arXiv preprint, arXiv:1804.03999, 2018.

[6] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen, “ResUNet++: An advanced architecture for medical image segmentation,” in 2019 IEEE International Symposium on Multimedia (ISM), 2019.

[7] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, “Dual attention network for scene segmentation,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[8] S. F. Banu, M. M. K. Sarker, M. Abdel-Nasser, D. Puig, and H. A. Raswan, “AWEU-Net: An attention-aware weight excitation U-Net for lung nodule segmentation,” Applied Sciences, vol. 11, no. 10132, 2021.

[9] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, and Y. Wang, “TransUNet: Transformers make strong encoders for medical image segmentation,” arXiv preprint, arXiv:2102.04306, 2021.

[10] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, “Swin-Unet: Unet-like pure transformer for medical image segmentation,” arXiv preprint, arXiv:2105.05537, 2021.

[11] LUNA16 Homepage, [Online]. Available: https://luna16.grand-challenge.org, last accessed Dec. 10, 2024.

[12] G. Tong, Y. Li, H. Chen, Q. Zhang, and H. Jiang, "Improved U-Net network for pulmonary nodules segmentation," Optik, vol. 174, pp. 460–469, 2018.

[13] N. V. Keetha and C. Sekhara, “U-Det: A modified U-Net architecture with bidirectional feature network for lung nodule segmentation,” arXiv preprint, arXiv:2008.03798, 2020.

[14] Z. Wu, Q. Zhou, and F. Wang, “Coarse-to-fine lung nodule segmentation in CT images with image enhancement and dual-branch network,” IEEE Access, vol. 9, pp. 7255–7262, 2021.

[15] M. Maqsood, S. Yasmin, I. Mehmood, M. Bukhari, and M. Kim, “An efficient DA-Net architecture for lung nodule segmentation,” Mathematics, vol. 9, no. 1457, 2021.

[16] C. Sekhara, N. V. Keetha, P. K. Donta, and G. Rajita, “A Bi-FPN-based encoder–decoder model for lung nodule image segmentation,” Diagnostics, vol. 13, no. 8, pp. 1406–1406, 2023.