ITC 4/54

Information Technology and Control
Vol. 54 / No. 4/ 2025
pp. 1326-1342
DOI 10.5755/j01.itc.54.4.39602

EMCA-UNet: A Novel Multi-Scale Convolutional Attention Model for Enhanced Lung Nodule Segmentation

Received 2024/11/27 | Accepted after revision 2025/08/11

HOW TO CITE: Han, L., Zheng, Z., Huang, Y., Zhang, H. (2025). EMCA-UNet: A Novel Multi-Scale Convolutional Attention Model for Enhanced Lung Nodule Segmentation. *Information Technology and Control*, 54(4), 1326-1342. https://doi.org/10.5755/j01.itc.54.4.39602

EMCA-UNet: A Novel Multi-Scale Convolutional Attention Model for Enhanced Lung Nodule Segmentation

Liying Han*, Zelin Zheng
School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China

Yiran Huang
School of International Education, Hebei University of Technology, Tianjin 300401, China

Hong Zhang
School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China

Corresponding author: lizhi_han@hebut.edu.cn

Pulmonary nodule segmentation plays a crucial role in the early detection of lung cancer. However, existing detection methods fail to efficiently extract multi-scale features and precisely reconstruct nodule boundaries, especially when dealing with nodules of varying sizes, irregular shapes, and complex backgrounds. To address this challenge, we introduce a novel U-Net architecture called EMCA-UNet (Efficient Multi-scale Convolution Attention). Specifically, the encoder of EMCA-UNet comprises Residual Multi-scale Attention Convolution (RESMAC) blocks, which enhance the model's feature extraction capabilities. The decoder integrates three modules: Multi-scale Convolution Attention (EMCA), Large Kernel Attention Gates (LGAG), and Efficient Up-Convolution Blocks (EUCB). These modules synergistically form a new multi-scale attention decoding layer that replaces the traditional U-Net decoder structure, thus enabling efficient multi-scale feature fusion and precise boundary reconstruction. Research has demonstrated that EMCA-UNet outperforms traditional models on publicly available datasets, such as LIDC-IDRI and LNDb. The Dice coefficient improved from 0.9198 to 0.9304, and the IoU increased from 0.8727 to 0.8856. Experimental results demonstrate that the proposed method offers a novel perspective for pulmonary nodule segmentation.

KEYWORDS: Lung Nodule Segmentation, U-Net, Convolutional Attention, Multi-Scale Feature, LIDC-IDRI.

1. Introduction

In recent years, both the incidence and fatality rates from lung cancer have gradually increased globally. According to data, the number of newly diagnosed lung cancer cases increased from 2.1 million in 2018 to 2.48 million by 2022. Lung cancer is one of the top causes of cancer-related fatalities worldwide [3, 28, 4]. The high death rate of lung cancer is largely due to the absence of visible early signs, with the majority of patients being detected at an advanced stage. Early identification and treatment of lung cancer can drastically lower death rates. Medical studies show that regular LDCT screening for pulmonary nodules in high-risk groups can effectively lower lung cancer mortality [1]. The extensive volume of images generated by LDCT scans makes it challenging for radiologists to rapidly and accurately analyze all images. To solve this problem, CAD technology that automatically segments and locates pulmonary nodules has been introduced. This technology can significantly improve diagnostic accuracy, reduce radiologists' workload, and offer crucial support for global early lung cancer screening initiatives.

Accurate segmentation of pulmonary nodules in LDCT images serves as a foundational guarantee for subsequent diagnosis and treatment. However, this continues to be a challenging issue. Pulmonary nodules show substantial variability in shape and size, frequently accompanied by complex backgrounds. The intricate structure of the lungs leads to the adhesion or partial occlusion of nodules and background structures (e.g., blood vessels, trachea, and pleura). These structures hinder the precise detection and segmentation of pulmonary nodules. To address these issues, two major approaches are now being used: classical image processing and deep learning-based image processing algorithms. Traditional image segmentation methods include threshold-based segmentation, region-based edge detection, level-set approaches, and morphological segmentation. These methods are capable of handling nodules of varying sizes and shapes; however, they are highly sensitive to imaging noise and tissue contrast variations, making them less effective for complexly shaped nodules. In recent years, deep learning systems like as U-Net, U-Net++, and TransUNet have achieved substantial advances in medical picture segmentation, becoming industry standards. The U-shaped design has been widely used in medical picture segmentation, and numerous researchers are constantly enhancing it. For instance, U-Net++ and Attention U-Net optimize skip connections, while ResUNet integrates residual structures into the convolutional base module. TransUNet and Swin-UNet have also introduced additional innovations. TransUNet combines the local feature extraction powers of convolutional neural networks with the global contextual power of transformers, providing improved performance in multi-organ segmentation by effectively capturing long-range relationships. Swin-UNet, built on the Swin Transformer, combines the benefits of local feature extraction and multi-scale feature capture, improving the model's capacity to handle complicated and fine-grained structures. Despite these advancements, pulmonary nodule segmentation remains challenging due to the variability in nodule shapes and sizes, as well as the complex background.

The feature representation in EMCA-UNet is further enhanced through the integration of a global multi-scale convolutional attention mechanism. The network's encoder utilizes the RESMAC architecture, incorporating both EMA and ConvNeXt modules, which improve multi-scale feature extraction and accelerate convergence. EMA ensures stability during the convergence process, while ConvNeXt optimizes convolution operations and feature map handling, thereby enhancing efficiency. This model demonstrates superior performance in processing complex, noisy medical images, particularly in small nodule detection, and is capable of capturing fine details across multiple scales. The EMCA module in the decoder combines multi-scale features to accurately delineate complex boundaries, thereby enhancing segmentation precision. The LGAG module integrates local details with global semantic information, reducing noise while preserving high-resolution feature representation. The EUCB module extracts fine-grained features through up-convolution, efficiently reconstructing small nodules and their associated features, optimizing computational performance, and making the model well-suited for real-time clinical segmentation. Together, these modules address key challenges in small nodule detection, boundary delineation, and noise reduction, ultimately enhancing segmentation accuracy. Compared to other methods, EMCA-UNet proves to be more effective in detecting lung nodules.

The main contributions and innovations of this paper are as follows:

1. A new decoding layer structure is proposed consisting of modules EMCA LGAG and EUCB which replaces the traditional U-Net decoding layer. This design enables efficient multi-scale feature merging and improved boundary reconstruction.
2. A novel encoding module, RESMAC is developed as a new backbone network to address gradient vanishing enhance multi-scale feature extraction and accelerate model convergence.
3. A combined loss function is introduced to stabilize foreground and background features adaptively balancing in the model significantly improving its performance on complex medical images.

2. Related Work

Farag et al. [8] suggested a variational level set approach for pulmonary nodule segmentation that could manage nodules with different shapes, sizes, and locations in the anatomy. Kuhnigk et al. [15] performed an experiment where they employed a morphology-guided automatic pulmonary nodule segmentation, which was of varying size and encompassed those with complex pulmonary morphology. Mao et al. [18] have suggested a hybrid approach that applies window functions, Fourier filtering, and fuzzy C-means for nodule searching. Yang et al. [29] have redefined pixel grayscale values in the spatial domain and chosen suitable fuzzy factors for nodule segmentation. Bruntha et al. [5], with the use of suitable seed points and growth conditions, were able to segment nodules effectively. Keshani et al. [12] created a sophisticated system with a support vector machine (SVM) classifier, which operates in tandem with an active contour model to efficiently segment and detect nodules. The active contour model extracts nodule boundaries from CT images, while the SVM classifier is used for nodule identification and classification. Govindaraju et al. [26] used the EK-means clustering technique to separate pulmonary nodules in CT images with tumor detection and categorization. Kostis et al. [13] created a 3D segmentation method using spiral CT images to locate nodules and quantify growth, allowing doctors to evaluate the malignant potential of nodules. Although these methods have made significant progress, conventional segmentation approaches have the propensity to need a large amount of human intervention, and repeated iterations could lead to over-segmentation or under-segmentation. Moreover, their high complexity in computation also renders them less competitive for clinical use. As a result, deep learning algorithms are the most advanced methods for segmenting pulmonary nodules today.

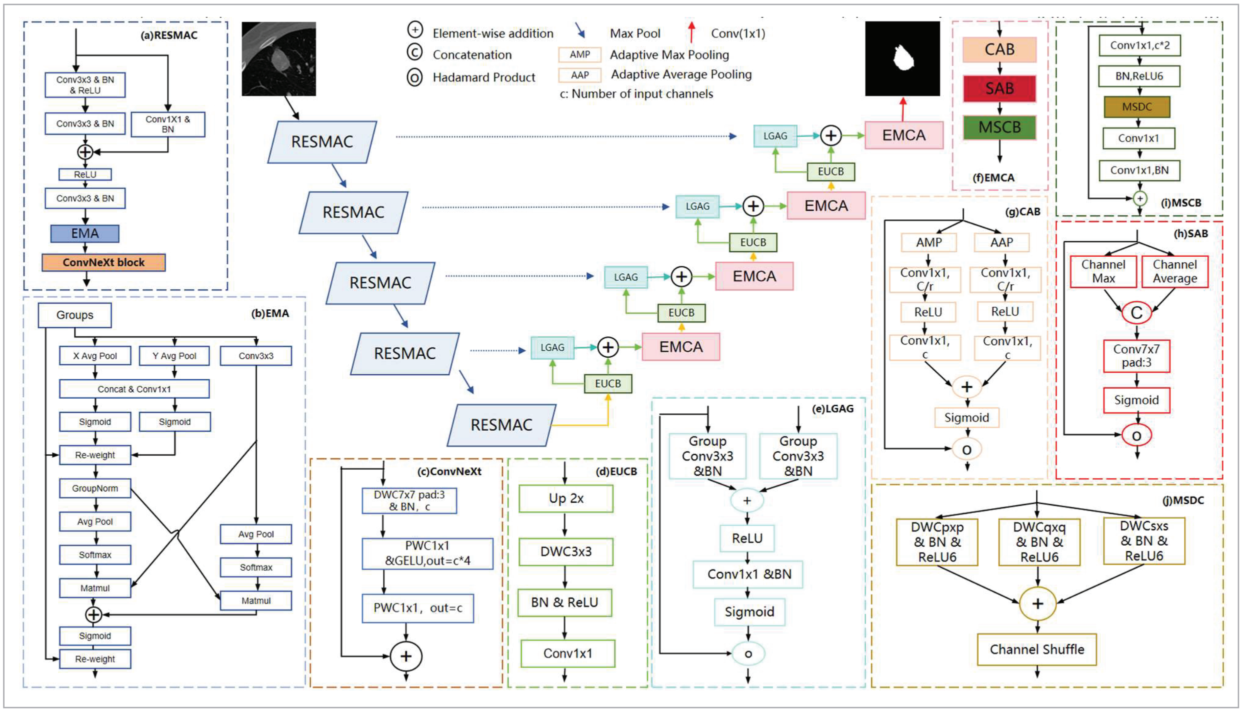
Since the introduction of AlexNet by Krizhevsky et al. [14], deep learning has achieved significant advances in image processing and altered medical image segmentation research. Ronneberger et al. [25] proposed U-Net, which is regarded as a seminal work in medical picture segmentation. U-Net uses skip connections to bypass features from the encoder to the decoder, maintaining fine-grained information in shallow layers while avoiding information loss due to pooling, resulting in maximum segmentation accuracy. Subsequent works have also kept expanding on various advances to the U-Net model to better solve more advanced medical image segmentation problems. Zhou et al. [30] presented U-Net++, an enhancement of multi-scale feature fusion by dense skip connections to learn more delicate morphological features and speed up model convergence. Alom et al. [2] proposed R2U-Net, which improves feature representation by employing recursively residual convolutional layers, establishing a good balance between accuracy and computation cost. Oktay et al. [19] presented Attention U-Net, which used attention gating modules to automatically eliminate unnecessary regions while focusing on target characteristics, particularly in complicated backdrops. Guan et al. [9] presented Dense U-Net, which allows for the reuse of features and the transmission of gradients via dense connections, so effectively reducing the vanishing gradient problem. Zhang et al. [21] introduced an improved U-Net3+ with stage residual for brain tumor segmentation. Schlemper et al. [27] introduced the AttentionResUNet. This model integrates attention mechanism and residual network for better adaptive attention to salient features for improved segmentation accuracy. Chen et al. [6] Integrated the local feature learning ability of convolutional neural networks and global contextual ability of transformers for better performance in multi-organ segmentation proposed TransUNet. Liu et al. [16] introduced the Swin-UNet, an extension of the Swin Transformer model. The extension well leverages the advantages of local feature extraction and multi-scale feature aggregation, hence making the computation more efficient and enabling more coherent representations. Saidi Guo et al. [7] proposed a causal knowledge fusion (CKF) framework that removes modal factors through causal intervention, extracts anatomical factors, and enhances the accuracy and information fusion capabilities of cross-modality cardiac image segmentation by incorporating a 3D hierarchical attention mechanism. Fallahpoor et al. [10] developed a deep learning-based model for liver and liver lesion segmentation, achieving high-precision liver segmentation on clinical 3D MRI data. Ramamoorthy et al. [24] integrated the Transformer and attention mechanisms into the U-Net structure and proposed TransAttU-Net to address the problem of accurate brain tumor segmentation. They [23] also combined the honey bee optimization algorithm with Probabilistic U-RSNet to construct a hybrid artificial intelligence method, which efficiently solves the problem of multi-class brain tumor image classification.

Despite the advances achieved by models such as U-Net, Attention U-Net, and Swin-UNet, limitations remain in accurately segmenting small nodules, addressing complex anatomical variations, and preserving high-resolution details.

In this paper, EMCA-UNet is introduced, integrating a multi-scale convolutional attention mechanism and an efficient up-convolution module, delivering enhanced feature extraction and boundary reconstruction capabilities. Unlike prior models such as Attention U-Net and TransUNet, which primarily focus on incorporating attention mechanisms, EMCA-UNet enhances segmentation accuracy by integrating residual multi-scale attention with efficient multi-scale convolutional attention decoding in the encoder. This allows EMCA-UNet to outperform these prior models in accurately segmenting small nodules within complex anatomical backgrounds, offering a more dependable solution for clinical applications.

Figure 1

The overall structure of EMCA-UNet is described as follows. The encoding layer, (a) RESMAC, consists of (b) EMA and (c) ConvNeXt. In the decoding layer, (d) EUCB is responsible for upsampling, (e) LGAG processes skip connection information, and (f) EMCA further recovers detailed information. Component (f) is composed of (g), (h), (i), and (j).



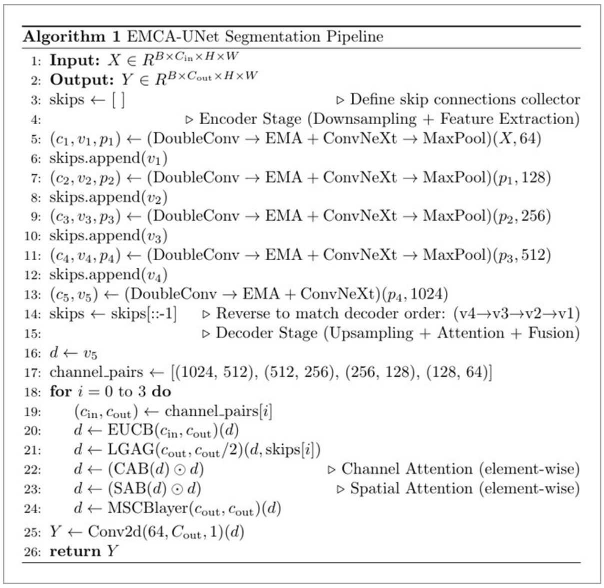
3. Methods

3.1. EMCA-UNet General Architecture

The EMCA-UNet model presented in this paper markedly improves pulmonary nodule segmentation performance. At its core, the EMCA-UNet encoder incorporates the Residual Multi-scale Attention Convolution (RESMAC) module, which seamlessly integrates large convolution kernels with a multi-scale attention mechanism. This design enables the model to capture tm.,he edge details of pulmonary nodules across multiple dimensions (horizontal, vertical, and global), thereby facilitating the effective separation of the nodules from complex backgrounds. To improve the capacity of the model to recover nodule details and accurately reconstruct small nodule boundaries, we set up the Enhanced Multi-scale Convolutional Attention (EMCA) module, Large Kernel Group Attention Gating (LGAG), and Efficient Up-Convolution (EUCB) modules. These modules are seamlessly integrated into a novel multi-scale attention decoding layer. These modules enable the model to interpret bottleneck layer information effectively and progressively fuse data from different levels of the encoder, thereby leading to more precise nodule boundary delineation. The overall architecture is depicted in Figure 1. The algorithm flow is shown in Figure 2. The model collects multi-scale characteristics from the input CT images and, using a succession of encoding, decoding, and attention modules, achieves an accurate pulmonary nodule segmentation.

Figure 2

Algorithm flow: The algorithm flow of EMCA-UNet.

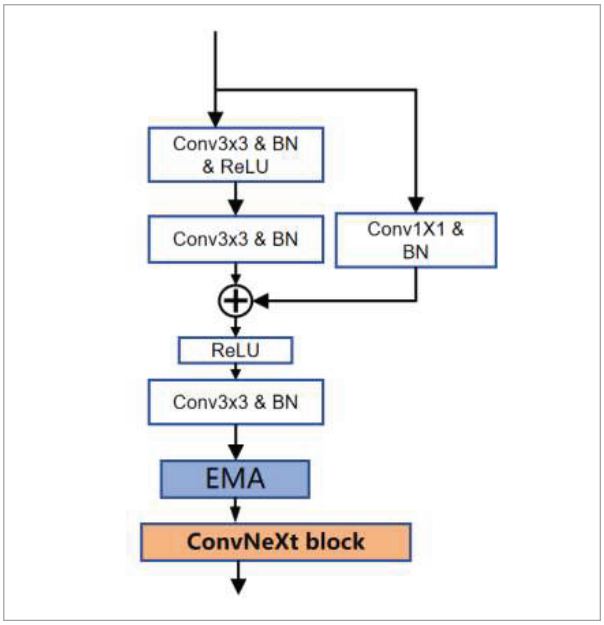


```
Algorithm 1 EMCA-UNet Segmentation Pipeline
Input: $X \in R^{B \times C_{\text{in}} \times H \times W}$
Output: $Y \in R^{B \times C_{\text{out}} \times H \times W}$
skips ← [ ]                                   ▷ Define skip connections collector
                         ▷ Encoder Stage (Downsampling + Feature Extraction)
$(c_1, v_1, p_1) \leftarrow$ (DoubleConv → EMA + ConvNeXt → MaxPool)$(X, 64)$
skips.append($v_1$)
$(c_2, v_2, p_2) \leftarrow$ (DoubleConv → EMA + ConvNeXt → MaxPool)$(p_1, 128)$
skips.append($v_2$)
$(c_3, v_3, p_3) \leftarrow$ (DoubleConv → EMA + ConvNeXt → MaxPool)$(p_2, 256)$
skips.append($v_3$)
$(c_4, v_4, p_4) \leftarrow$ (DoubleConv → EMA + ConvNeXt → MaxPool)$(p_3, 512)$
skips.append($v_4$)
$(c_5, v_5) \leftarrow$ (DoubleConv → EMA + ConvNeXt)$(p_4, 1024)$
skips ← skips[::-1]    ▷ Reverse to match decoder order: (v4→v3→v2→v1)
                       ▷ Decoder Stage (Upsampling + Attention + Fusion)
$d \leftarrow v_5$
channel_pairs ← [(1024, 512), (512, 256), (256, 128), (128, 64)]
for $i = 0$ to 3 do
    $(c_{\text{in}}, c_{\text{out}}) \leftarrow$ channel_pairs[$i$]
    $d \leftarrow \text{EUCB}(c_{\text{in}}, c_{\text{out}})(d)$
    $d \leftarrow \text{LGAG}(c_{\text{out}}, c_{\text{out}}/2)(d, \text{skips}[i])$
    $d \leftarrow (\text{CAB}(d) \odot d)$              ▷ Channel Attention (element-wise)
    $d \leftarrow (\text{SAB}(d) \odot d)$              ▷ Spatial Attention (element-wise)
    $d \leftarrow \text{MSCBlayer}(c_{\text{out}}, c_{\text{out}})(d)$
$Y \leftarrow \text{Conv2d}(64, C_{\text{out}}, 1)(d)$
return $Y$
```

3.2. Structure of RESMAC

Figure 3

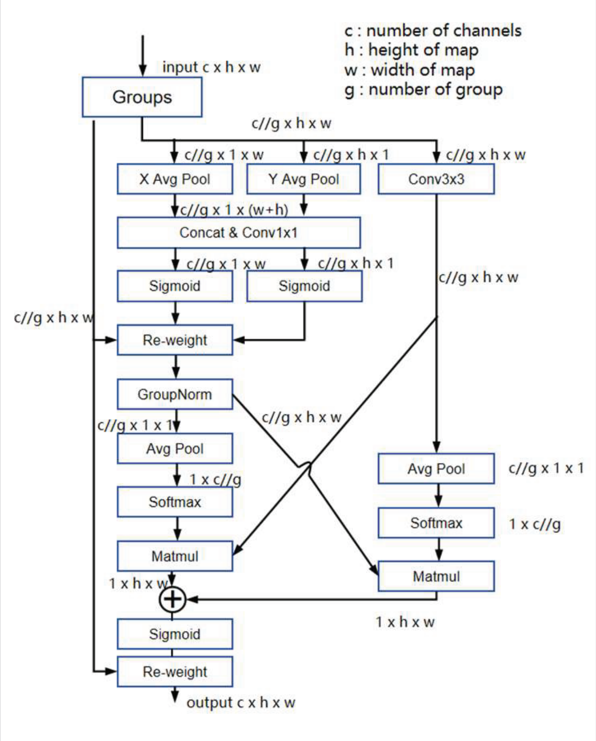
Structure of RESMAC: The module consists of a residual network, EMA, and ConvNeXt.



The new core module of the encoder, RESMAC, is depicted in Figure 3. This module integrates the Efficient Multi-scale Attention (EMA) residual convolution module with the large kernel ConvNeXt module, significantly boosting the model's capacity to detect and highlight critical pulmonary nodule features. The RESMAC module optimizes the feature extraction process, enhances the overall performance of the network, and accelerates model convergence. As a key component of RESMAC, the EMA module enhances feature-capturing capability through multi-scale adaptive pooling and inter-channel weighting, thereby enabling the model to acquire rich and critical feature representations at early stages. This module is particularly well-suited for capturing edge details of pulmonary nodules, effectively reducing background noise, and emphasizing nodule features at multiple scales, which is critical for medical image segmentation.

Figure 4

Structure of EMA: 'g' represents groups, 'X Avg Pool' denotes 1D global average pooling along the horizontal axis, and 'Y Avg Pool' denotes 1D global average pooling along the vertical axis.



In the RESMAC structure, illustrated in Figure 4, the EMA module [17] is placed after the residual convolution layer. It processes the output feature map from the residual convolution, which has the shape (batch_size, channels, height, width), and divides the channel dimension into 32 groups. Each group undergoes adaptive pooling along the height and width to capture multi-scale global information, generating key features, particularly in the vertical and horizontal directions. Its horizontal and vertical characteristic decompositions are shown in formulas (1) and (2), respectively. This horizontal-vertical decomposition mechanism is analogous to the contour perception of images in human vision.

$$x_h=\frac{1}{W}\sum_{j=1}^{W} x(j) \quad (1)$$

$$x_w=\frac{1}{H}\sum_{i=1}^{H} x(i)\,. \quad (2)$$

The processed features are fused using 1x1 convolutions to form a multi-scale fused feature map. This feature map is further split, and channel weighting is applied through a sigmoid activation function, enhancing the focus on critical channels. The implementation process is shown in Formula (3). The channel weighting successfully enhances boundary features, enabling EMA to exhibit good sensitivity to the boundaries between nodules and blood vessels.

$$x_{hw}=\sigma\left(W_{1\times1}\times Concat(x_h,x_w)\right). \quad (3)$$

Subsequently, GroupNorm is applied to improve the spatial representation of nodule details. The attention weights are generated from the globally pooled features to assess the importance of both global and local information for each channel group. The 2D pooling formula consists of X and Y:

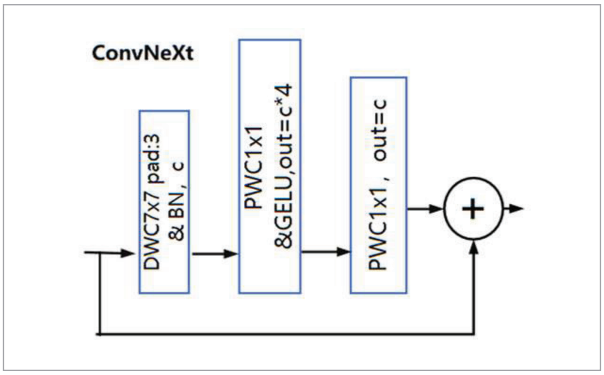
$$z_c=\frac{1}{H\times W}\sum_{i}^{H}\sum_{j}^{W} x_c(i,j) \quad (4)$$

The feature values are adjusted by matrix multiplication and reshaped to produce the output feature map which has the shape (batch_size, channels height width). The final implementation of EMA is shown in Formula (5).

$$y=GroupConv\left(x\times x_{hw}\times Softmax(z_c)\right) \quad (5)$$

Figure 5

Structure of ConvNeXt: After extracting more comprehensive information through large kernel convolutions, the channels are adjusted using pointwise convolutions with the input and output channel numbers remaining unchanged.



The Multi-scale Pooling, the adaptive channel weighting and convolution operations in the EMA module efficiently integrate global and local information. This enables the model to emphasize critical features in complex backgrounds thereby improving the accuracy of pulmonary nodule segmentation

As illustrated in Figure 5, the ConvNeXt module employs 7x7 depthwise separable convolutions and 1x1 pointwise convolutions combined with the GELU activation function to enhance the model's feature-capturing capabilities while maintaining efficiency. The design of this large-kernel convolution expands the model's receptive field, enabling it to capture the spatial relationships between nodules and their surrounding environment (such as the distance from the thoracic cavity). Not only has this design reduced computational complexity but it also improves the model's ability to segment complex scenes particularly when handling nodules of different sizes and shapes. Inspired by the Transformer architecture, ConvNeXt improves computational efficiency and feature representation optimizing high precision and speed in medical image segmentation. The specific formula is as follows:

$$\mathrm{GELU}(x)=x\times\frac{1}{2}\left[1+\frac{2}{\sqrt{\pi}}\int_{0}^{\frac{x}{\sqrt{2}}}e^{-t^{2}}\,dt\right] \tag{6}$$

$$y=C_{1\times1}\{\mathrm{GELU}[C_{1\times1}(\mathrm{BN}(\mathrm{DWC}_{7\times7}(x)))]\}+x\,. \tag{7}$$

Overall integration of EMA and ConvNeXt in the module RESMAC gives high-quality input features for the subsequent modules EMCA LGAG and EUCB providing significantly improved segmentation accuracy at challenging pulmonary nodules regions. This integrated design effectively balances computational efficiency with robust feature representation, enabling EMCA-UNet to achieve both high accuracy and rapid processing in medical image applications.

3.3. Multi-scale Attention Fusion Decoding Layer

In the EMCA-UNet decoder structure, traditional up-sampling is replaced by the Efficient Up-Convolution Block (EUCB). The LGAG (Local-Global Attention with Grouping) module integrates high-resolution skip connection features with up-sampled decoding features via grouped convolution. The Enhanced Multi-scale Convolutional Attention (EMCA) module serves as the core of the decoder, receiving combined inputs from the EUCB and LGAG modules and fusing multi-scale information to provide robust support for the fine-grained reconstruction of pulmonary nodule features. The configuration of the decoder layer is depicted in Figure 6.

The bottleneck layer and data from the preceding encoder layer are passed into the EUCB module. Through up-convolution operations with a ratio of 2, the spatial resolution of the input feature map is gradually improved to match the feature map of the subsequent skip connection. Following up-sampling, the feature map is then fine-tuned with 3x3 depthwise convolutions and maintain spatial structure without any additional computational cost. The pulmonary nodule feature map obtained is then batch normalized and activated by ReLU to ensure uniform feature distribution across scales as well as speeding up model convergence. In addition, 1x1 convolution is employed to alter the amount of channels so that the map of features integrates well with the following layer. The formula is illustrated in detail as follows:

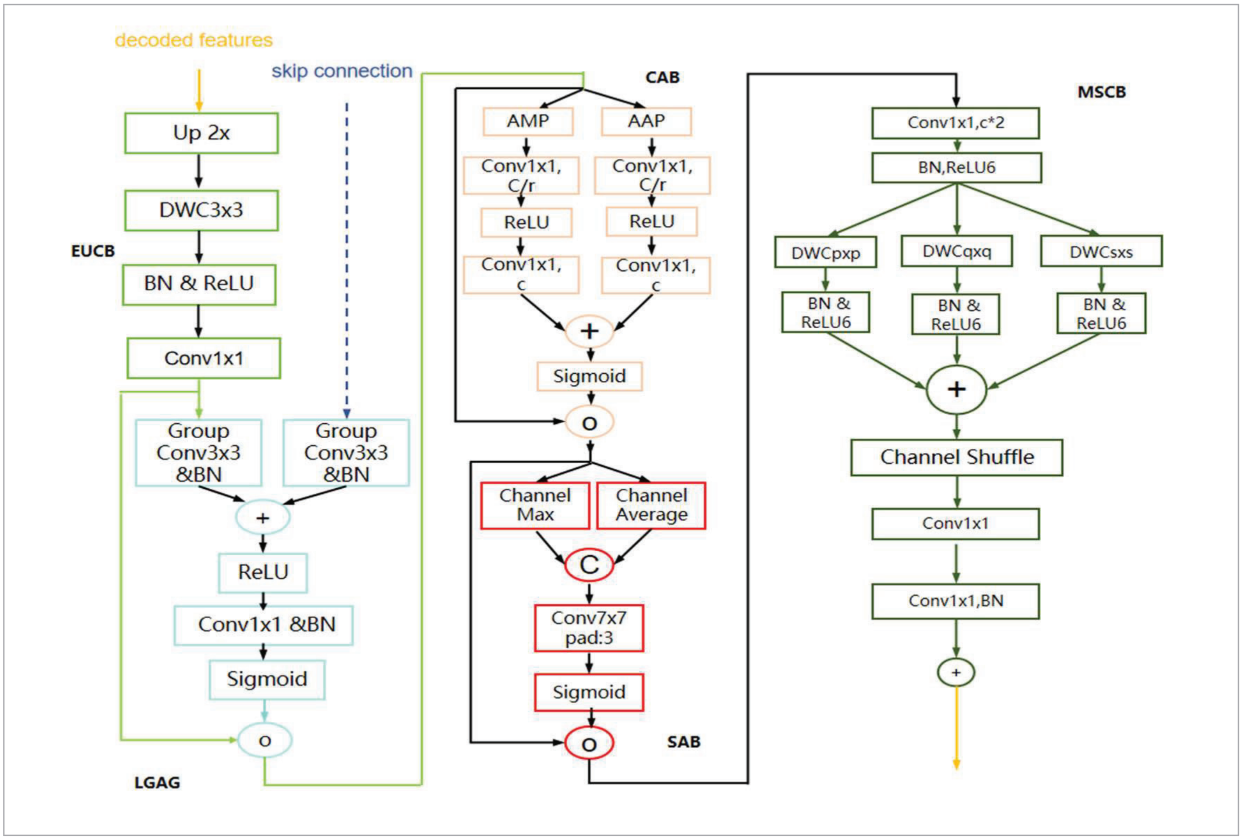
$$\mathrm{EUCB}(x)=c_{1\times1}\left\{\mathrm{ReLU}\left[\mathrm{BN}\left(DWC_{3\times3}\left(Up_{2}(x)\right)\right)\right]\right\}. \tag{8}$$

The LGAG module is designed to boost the model's focus on pulmonary nodules by controlling the up-sampled feature map along with the gating signal (from skip connection features) via 3x3 grouped convolutions (group=1). The features are merged through element-wise summation of high-resolution skip connection features and up-sampled decoding features. The resulting composite feature map is then processed through batch normalization and ReLU activation, thus boosting the representation of non-linear features. Following 1x1 convolution and batch normalization, attention coefficients produced by the sigmoid function are used to further modulate feature responses. Finally, the input feature x is scaled by element-wise multiplication. The specific formula is as follows:

$$q_{att}(g,x)=\mathrm{Relu}\left\{\mathrm{BN}\left[\mathrm{GC}_{g3\times3}(g)\right]+\mathrm{BN}\left[\mathrm{GC}_{x3\times3}(x)\right]\right\} \tag{9}$$

Figure 6

Structure of the Decoding Layer: Efficient Up-Convolution Block (EUCB), Multi-scale Convolutional Attention Module (MSCAM), Multi-scale Convolution Block (MSCB), Multi-scale (Parallel) Depth-Wise Convolution (MSDC), Large-Kernel Grouped Attention Gate (LGAG), Channel Attention Block (CAB), and Spatial Attention Block (SAB).



$$\mathrm{LGAG}(g,x)=x \circledast \sigma\left\{\mathrm{BN}\left[C_{1\times 1}\left(q_{att}(g,x)\right)\right]\right\}. \quad (10)$$

The EMCA module fuses multi-scale information during decoding, offering robust contextual guidance for feature recovery. This module is built on a multi-level convolutional attention design, incorporating Channel Attention Block (CAB) and Spatial Attention Block (SAB) modules, which apply adaptive weighting to both channels and spatial regions, enabling the decoder to focus on pulmonary nodule features at different scales, thereby enhancing segmentation precision. The specific formula is as follows:

$$\mathrm{EMCA}(x)=\mathrm{MSCB}\left(\mathrm{SAB}\left(\mathrm{CAB}(x)\right)\right). \quad (11)$$

The CAB module first performs adaptive max pooling (AMP) and adaptive average pooling (AAP) on the input feature map to extract global information. AMP captures the highest activation values in the feature map, focusing on prominent nodule regions, while AAP computes the average response, retaining overall channel information to balance AMP's focused effect. In this way, CAB assigns high weights to the CT value distribution specific to nodules, screening for the characteristic density of lesions. The outputs from AMP and AAP are reduced to one-sixteenth of the original channel count via a 1x1 convolution to decrease computational complexity, followed by ReLU activation to enhance non-linearity. The reduced features are restored to the original channel count using another 1x1 convolution. The fused feature map is created by merging the AMP and AAP data and then using a sigmoid activation function that is used to calculate the channel attention weights. In this approach, the CAB module weights each channel adaptively, highlighting critical information while reducing redundancy.

$$\mathrm{CAB(x)}=\sigma\{C_{1\times1}[\mathrm{Relu}(C_{1\times1}(P_m(x)))] +C_{1\times1}[\mathrm{Relu}(C_{1\times1}(P_a(x)))]\}\circledast x. \tag{12}$$

The SAB module further focuses on key regions in the spatial dimension, ensuring precise capture of boundary information for pulmonary nodules. Utilizing channel max pooling and channel average pooling, the SAB module generates two single-channel feature maps that represent the maximum and average responses across channels for each pixel. Max pooling emphasizes salient regions, while average pooling softens the background. The two feature maps are added along the channel axis to create a two-channel feature map. This is subjected to a 7x7 convolution to facilitate spatial feature extraction, hence enhancing the boundary perception quality of the pulmonary nodule. The input feature map is handled using a sigmoid function to generate spatial attention weights, which are then used to influence the input map's spatial features. This works to highlight the areas where the nodules are present and minimizes extraneous background noise. Simultaneously, the combination of 7×7 convolution and channel pooling can concentrate attention on the contour details of nodules, assigning greater weights to boundary regions with significant curvature changes in the segmentation of irregular nodules. The architecture enables the SAB module to focus on key locations in the spatial domain, enhancing the model's ability to highlight the contours and forms of pulmonary nodules during segmenting.

$$\mathrm{SAB(x)}=\sigma\left\{C_{7\times7,pad=3}\left[Ch_{max}(x),Ch_{avg}(x)\right]\right\}\circledast x. \tag{13}$$

The MSCB module is responsible for the acquisition of multiscale pulmonary nodule features necessary for processing nodules of different sizes and shapes in compound medical images. The MSCB module acquires features by using 1x1, 3x3, and 5x5 depthwise separable convolutions (DWC) to harvest information at different scales. The features from different convolutions are concatenated or summed along the channel axis and shuffled randomly to achieve uniform distribution across spatial and channel axes. The compound feature map undergoes pointwise convolution for combination and compression, yielding the output features. The multi-scale convolution can be represented mathematically as:

$$\mathrm{MSDC(x)}=\sum_{ks\in KS}\mathrm{DWCB}_{ks}(x). \tag{14}$$

In this context, ks represents different kernel sizes (e.g., 1, 3, 5). The complete formula of MSCB is as follows.

$$\mathrm{MSCB(x)}=\mathrm{BN}\{C_{1\times1}[\mathrm{CS}(\mathrm{MSDC}(\mathrm{R6}(\mathrm{BN}(C_{1\times1}(x)))))]. \tag{15}$$

The EMCA module effectively fuses multi-scale features, enhancing the efficiency of the decoding process and improving segmentation accuracy. This design improves the ability to capture complex features, particularly in medical image segmentation, where fine-grained details and boundaries of pulmonary nodules must be accurately delineated.

3.4. Loss Function

In medical picture segmentation, the Dice Similarity Coefficient (DSC) and Binary Cross-Entropy (BCE) losses are commonly used. In segmentation tasks, pulmonary nodules are very tiny compared to the backdrop, resulting in a large class imbalance between positive and negative samples. As a result, a mix of Dice and BCE loss was used to balance the learning of local details with global frameworks.

The Dice loss is effective in addressing the foreground-background imbalance by focusing on the foreground area, encouraging the model to prioritize the nodule region. However, Dice loss can produce unstable gradients when the target area is small. Conversely, BCE loss is well-suited for pixel-wise binary classification, providing precise gradient information during the initial stages of segmentation. The definitions of the Dice and BCE loss functions are as follows:

$$L_{dice}=1-\frac{2|X\cap Y|}{|X|+|Y|} \tag{16}$$

$$L_{BCE}=-\frac{1}{N}\sum_{n=1}^{N}W_n\left[y_n\log x_n+(1-y_n)\log(1-x_n)\right]. \tag{17}$$

In this context, |X∩Y| represents the intersection of the predicted segmentation region and the ground truth mask, while |X|+|Y| denotes the total number of elements in both. The hyperparameter w controls the weight of each pixel, and y is the ground truth label (0 or 1).

To balance the overall structure and fine details, this paper uses a combined loss function:

$$L=0.5\times L_{BCE}+L_{dice}. \quad (18)$$

The BCE loss guides the Dice loss at the micro-level, providing stability during learning on imbalanced data. By combining BCE and Dice losses, the network receives more stable gradients during the early stages, preventing the gradient vanishing phenomenon that may occur initially with Dice loss. This approach enables the model to more accurately locate the lung nodule regions.

4. Experimental

4.1. Datasets

Two widely used public lung nodule datasets, LIDC-IDRI and LNDb, were employed. These datasets were selected for their richness, generalizability, universality, and high-quality annotation standards, providing a reliable foundation for lung nodule segmentation experiments. The LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) dataset, initiated by the U.S. National Cancer Institute, contains CT images from 1,018 patients, each annotated independently by four radiologists. To fully leverage the annotations from each expert, a logical OR operation was applied (considering a region as a nodule if at least one radiologist labeled it as such), resulting in the creation of new mask images.

The LNDb dataset includes 294 lung CT scans retrospectively collected from the Centro Hospitalar Universitário de São João (CHUSJ) in Porto, Portugal, between 2016 and 2018, with detailed annotations provided by radiologists. For both datasets, the annotations were processed by extracting slices containing nodules and their corresponding masks, then cropping the masks to a 128×128 resolution centered on the pulmonary nodules and saving them as PNG images. Due to variations in scanners and environmental factors, the quality of CT images can vary. Additionally, annotations for both datasets were provided by multiple radiologists, each following different standards, which necessitated the merging of annotations using a logical OR operation. This approach may increase the difficulty of data segmentation. Moreover, the relatively low resolution of the images in both datasets limits feature extraction, particularly for small nodules, thereby reducing segmentation accuracy. Despite preprocessing, the complex shapes of the nodules and the presence of background noise continue to pose significant challenges.

The LIDC-IDRI dataset contains 15,096 images, which were split into training, validation, and test sets in a 60:20:20 ratio, resulting in 9,057 images (60%) for training, 3,019 images (20%) for validation, and 3,020 images (20%) for testing. The LNDb dataset contains 490 images, divided in an 80:10:10 ratio, yielding 392 images (80%) for training, 49 images (10%) for validation, and 49 images (10%) for testing. Preprocessing for both datasets includes grayscale normalization, resolution standardization, and cropping around nodules to ensure consistency in model input. In the paper, a fixed data partition is adopted. For both LIDC - IDRI and LNDb, the datasets are partitioned in fixed proportions to avoid random differences caused by different k - fold cross - validations.

4.2. Evaluation Indicators

To comprehensively assess model performance, five metrics are employed: Dice, IoU, Precision, Recall, F1-Score, and MCC. In lung nodule segmentation tasks, Dice and IoU are the core metrics, primarily used to measure the overlap between the segmented area of the model and the ground truth.

$$\text{Dice}=\frac{2\times TP}{(FP+TP)+(FN+TP)} \quad (19)$$

$$\text{IoU}=\frac{TP}{TP+FP+FN}. \quad (20)$$

Additionally, the model's performance on local details is evaluated using Precision, Recall, F1-Score, and MCC. The definitions of these metrics are as follows:

$$\text{Precision}=\frac{TP}{TP+FP} \quad (21)$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \tag{22}$$

$$\text{F1-Score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{23}$$

$$\text{MCC}=\frac{\text{TP}\times\text{TN-FP}\times\text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FN})(\text{TN+FP})(\text{TN+FN})}} \tag{24}$$

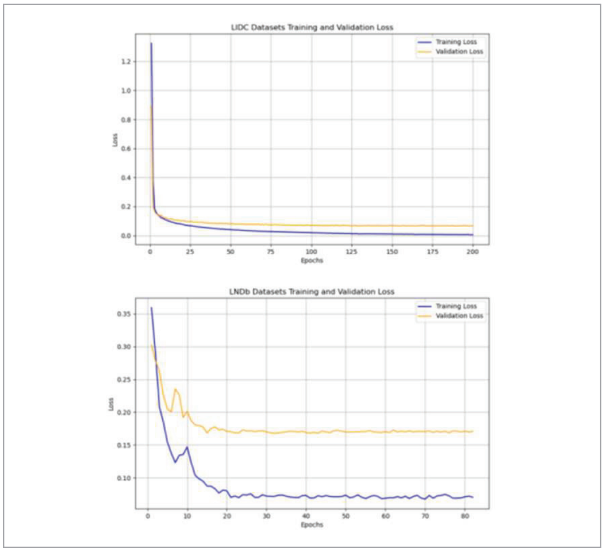
where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

4.3. Training Setup

In the experimental setup, during the training phase on the LIDC dataset, the model is configured with a batch size of 8, a maximum of 200 iterations, and the AdamW optimizer, with an initial learning rate of 1e-5. For the smaller LNDb dataset, to enhance training efficiency and performance, the learning rate is adjusted to 0.001, while maintaining a batch size of 8 and a maximum of 100 iterations. Additionally, a transfer learning strategy is employed, utilizing the model trained on the LIDC dataset as a pre-trained model for the LNDb dataset. This approach accelerates the training process while enhancing segmentation performance. An early stopping criterion is implemented after 50 iterations to prevent overfitting. In this configuration, the validation loss stabilized without further decrease, and neither overfitting nor underfitting was observed.

Figure 7

The loss curves for the LIDC and LNDb datasets during training.



As shown in Figure 7. To ensure the fairness of comparative experiments, all comparative experiments in this study were retrained on the same dataset and compared using the same evaluation metrics.

All experiments are conducted in a Python 3.9 and PyTorch 2.2.2 environment, on a computer with Ubuntu 20.04 and an RTX 4090 GPU.

5. Experimental Results and Analysis

5.1. Main Experiment

To evaluate the performance of the proposed model, it is compared to known segmentation models such as conventional U-Net and Attention U-Net. The studies use standard assessment measures including the dice coefficient, intersection over union (IoU), and accuracy. Table 1 shows that on the LIDC dataset, EMCA-UNet outperforms standard models like as U-Net, U-Net++, and Attention U-Net, with a Dice coefficient of 0.9304 and an IoU of 0.8856. Furthermore, EMCA-UNet achieves good precision and recall scores of 0.9451 and 0.9339, indicating a proportionate reduction in both false positives and false negatives. The significant improvements in Dice and IoU indicate the model's superior segmentation accuracy, while the high Precision and Recall values reflect its capability to accurately delineate target regions and effectively suppress background noise. These results suggest that the proposed model offers more precise segmentation of pulmonary nodules.

Figure 8 illustrates the segmentation results of several models on the LIDC dataset, including U-Net, U-Net++, Attention U-Net, TransUNet, SMRUNet [22], EMCADNet [11], and the proposed EMCA-UNet. As shown in Table 1, EMCA-UNet achieves Dice and IoU scores of 0.9304 and 0.8856, respectively, outperforming other models by a significant margin. The improvements in Dice and IoU reflect EMCA-UNet's superior segmentation accuracy, while the high Precision and Recall scores demonstrate its effectiveness in minimizing both false positives and false negatives. The visual results further illustrate that EMCA-UNet excels in capturing boundary details, especially in irregularly shaped nodule regions. For instance, in images (d) and (e), the boundary details of U-Net and other models appear relatively blurred, while EMCA-UNet precisely delineates the nodule contours, effectively avoiding both under-segmentation and over-segmentation. These findings indicate that EMCA-UNet maintains robust segmentation accuracy even in complex backgrounds and ambiguous boundary conditions, thereby enhancing its ability to accurately segment pulmonary nodule regions.

Table 1

Comparison of quantitative results of different models.

Method	Dice	IOU	Precision	Recall	F1	MCC
UNet	0.9198	0.8727	0.9339	0.9266	0.9201	0.9243
Unet++	0.9225	0.8767	0.9372	0.9280	0.9230	0.9270
AttentionUNet	0.9214	0.8750	0.9371	0.9248	0.9215	0.9250
TransUNet	0.9184	0.8700	0.9323	0.9256	0.9188	0.9244
SwinUNet	0.9150	0.8761	0.9341	0.9238	0.9200	0.9238
SMRUNet	0.9213	0.8734	0.9367	0.9266	0.9219	0.9297
EMCADNet	0.9143	0.8484	0.9293	0.9070	0.9164	0.9210
EMCA-UNet	**0.9304**	**0.8856**	**0.9451**	**0.9339**	**0.9311**	**0.9342**

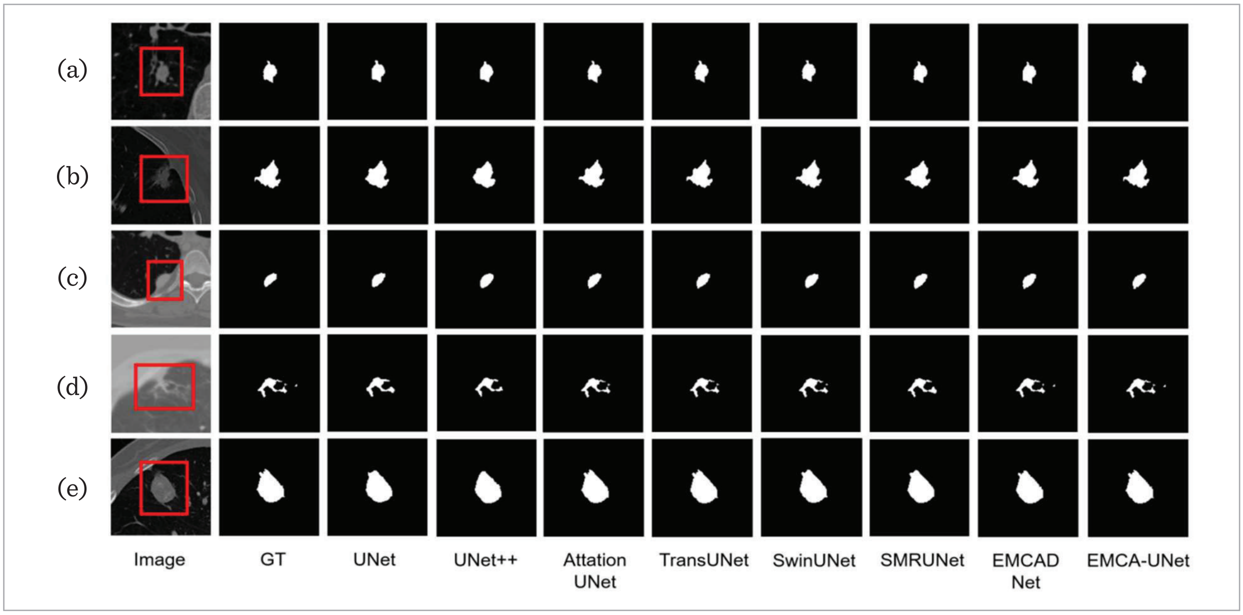
Table 2

Results of ablation experiments.

Method	Dice	IOU	Precision	Recall	F1	MCC
UNet	0.9198	0.8727	0.9339	0.9266	0.9201	0.9243
UNet+ConvNeXt	0.9236	0.8776	0.9381	0.9285	0.9243	0.9288
UNet+RESMAC(ConvNeXt+EMA)	0.9267	0.8828	0.9401	0.9318	0.9271	0.9307
UNet+RESMAC+EMCAD	**0.9304**	**0.8856**	**0.9451**	**0.9339**	**0.9311**	**0.9342**

Figure 8

The segmentation results for a typical sample, with each row including the original picture, the segmentation result of Standard U-Net, the segmentation result of Attention U-Net, and the segmentation result of the suggested approach.



5.2. Ablation Experiments

An ablation study was carried out to determine the contribution of each module to the model's performance. Modules such as the attention mechanism and feature improvement components were added gradually, and the ensuing performance changes were examined to determine the importance of each part within the broader framework. The results of the ablation investigation are reported in Table 2.

Figure 9

Comparison of segmentation effects of ablation.

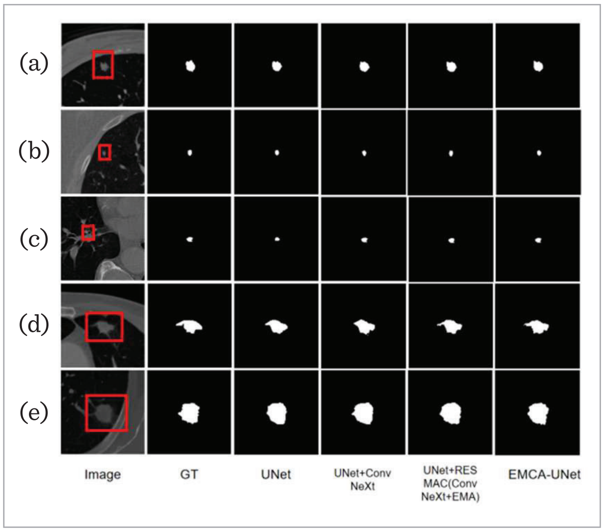


Figure 9 shows the findings of the ablation investigation, comparing the original U-Net with segmentation outcomes achieved by progressively integrating various modules, including ConvNeXt, RESMAC (a residual module combined with EMA), and ultimately the full EMCA-U-Net. Table 2 shows that each module contributes significantly to the overall model performance. For example, the RESMAC module increases the Dice coefficient by 0.41% and the IoU by 0.72%, while the addition of the EMCAD module further elevates the Dice and IoU scores to 0.9304 and 0.8856, respectively. The findings highlight the unique contribution of each module towards improving the U-Net architecture. The use of ConvNeXt is a significant factor in improving the ability of the U-Net to learn edge information. This research provides valuable insights into the complementary roles of the individual modules towards improving segmentation accuracy and model performance in general. This is mainly due to the inclusion of a large convolutional kernel in ConvNeXt architecture, which expands the receptive field and allows more extensive feature extraction. However, even with such improvements, segmentation performance remains constrained when handling certain difficult or occluded areas. The use of RESMAC, a cutting-edge multi-scale attention mechanism that combines the benefits of EMA and ConvNeXt, increases the model's capacity to accurately delineate pulmonary nodules in their local context. In addition, the use of multi-scale attention mechanisms within RESMAC enables the model to be more sensitive to nodules with various shapes and sizes, thus improving the precision and clarity of the resulting segmentation. The use of RESMAC significantly enhances the precision of identification and segmentation of nodules of complex shapes. As clearly seen in part (d) of the visual results, the baseline U-Net tends to lose or fail to recognize significant areas of the nodule, while the RESMAC-integrated model accurately detects and identifies these regions, thereby eliminating the issue of missed detection. Finally, EMCA-U-Net combines an advanced decoding process with multi-scale attention, achieving unprecedented advances in image detail restoration and handling complex shapes and boundary areas compared with the traditional U-Net. EMCA-U-Net effectively utilizes information at different scales to create a more sophisticated and richer context, thereby achieving better segmentation accuracy and more effective detail representation.

5.3. Validation of the LNDb Dataset

To test the suggested method's generalizability, it was implemented to the LNDb dataset and compared to the performance of the conventional U-Net and Attention U-Net. The experimental findings show that the suggested technique performs well on the LNDb dataset, demonstrating great generalizability.

Figure 10 shows the outcomes of the segmentation of several models on the LNDb dataset, demonstrating EMCA-UNet's capacity to handle complicated forms and nodules against tough backdrops. The LNDb dataset is characterized by small nodules and a complex background, posing a greater challenge for model generalization.

As shown in Table 3, EMCA-UNet achieves Dice and IoU scores of 0.8314 and 0.7196, respectively, on the LNDb dataset, outperforming other models and underscoring its superior generalization ability. This performance advantage can be attributed to EMCA-UNet's design optimizations, particularly in multi-scale attention and precise boundary capture. Intuitive visual comparisons show that EMCA-UNet clearly and accurately depicts nodule boundaries on the LNDb dataset, especially in regions with complex boundaries. For example, in comparisons (c) and (b), EMCA-UNet demonstrates a high degree of consistency with the Ground Truth (GT), accurately recognizing and delineating nodule regions without excessive smoothing. In contrast, traditional U-Net and Attention U-Net exhibit blurred boundaries and loss of detail in these regions, likely due to their limitations in handling highly complex backgrounds.

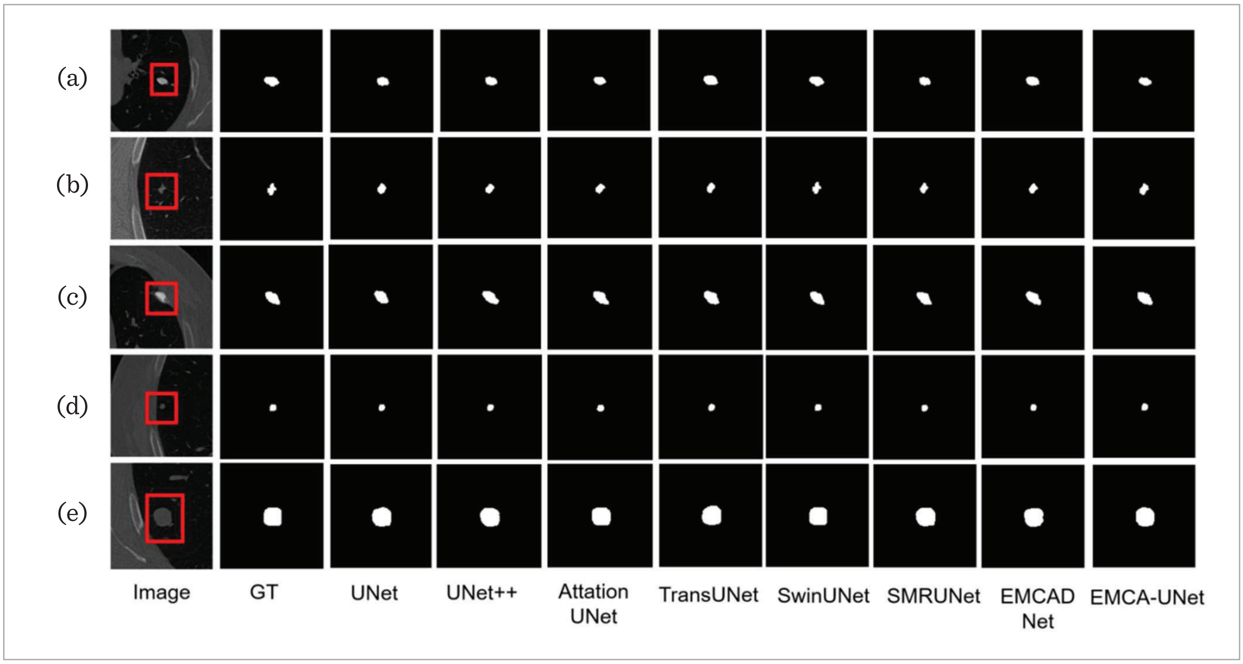
Table 3

Performance of different models on the LNDb dataset.

Method	Dice	IOU	Precision	Recall	F1	MCC
UNet	0.7953	0.6720	0.8886	0.7358	0.7955	0.7992
Unet++	0.7828	0.6646	0.8781	0.7235	0.7825	0.7861
AttentionUNet	0.8159	0.6970	0.8889	0.7713	0.8162	0.8196
TransUNet	0.8206	0.7066	0.8732	0.7888	0.8226	0.8265
SwinUNet	0.7961	0.6822	0.8804	0.7685	0.8024	0.8061
SMRUNet	0.8099	0.6871	0.8861	0.7477	0.8106	0.8143
EMCADNet	0.8056	0.6852	0.8425	0.7884	0.8085	0.8122
EMCA-UNet	**0.8314**	**0.7196**	**0.8729**	**0.8091**	**0.8329**	**0.8367**

Figure 10

Comparison of experimental results for the LNDb dataset.



6. Conclusion

The introduction of the EMCA-UNet model has led to notable enhancements in the fundamental U-Net architecture, achieving improved accuracy in pulmonary nodule segmentation. Through the integration of a multi-scale convolutional attention mechanism, the EMCA-UNet has been very effective at detecting small nodules and managing fuzzy boundaries. Importantly, on the leading LIDC-IDRI and LNDb public datasets for pulmonary nodule analysis, the model achieved outstanding performance in essential segmentation tasks, demonstrating good generalization ability. Moreover, EMCA-UNet significantly enhances the discrimination of complex pulmonary structures by dynamically fusing multi-level convolutional features and attention modules, which possesses great flexibility and clinical value in application.

Although EMCA-UNet has achieved positive results in multiple fields, there are still some limitations in practical applications. For example, when dealing with multiple overlapping pulmonary nodules or nodules with significant local scale differences, the segmentation performance of the model may be affected. The model may have difficulty accurately defining the contours of overlapping nodules, which may lead to focusing on larger nodules at the expense of smaller ones. In terms of parameter scale, the number of parameters of EMCA-UNet is 39 M, which significantly increases the complexity compared with the basic UNet architecture. However, compared with the improved models based on Transformer, it has obvious advantages in the number of parameters; compared with other improved schemes using attention mechanisms, the number of parameters is at the same order of magnitude. This characteristic of parameter scale restricts the deployment and operation of EMCA-UNet in application scenarios with limited computing resources. Developing lightweight models or optimization techniques can enhance its usability and clinical applicability. Future research should also explore human-AI collaboration to ensure the interpretability and transparency of model outputs, thereby enhancing the trust of clinicians and facilitating its clinical application. Research focused on improving model interpretability will ensure that artificial intelligence becomes a powerful support tool for healthcare professionals.

Acknowledgment

This work was supported by the Hebei Province introduction of overseas students funding project in 2022, “Research on lung CT image detection technology” (C20220316).

References

1. Aberle, D. R., Adams, A. M., National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. New England Journal of Medicine, 2011, 365(5), 395-409. https://doi.org/10.1056/NEJMoa1102873
2. Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., Asari, V. K. Recurrent Residual Convolutional Neural Network Based on U-Net (R2U-Net) for Medical Image Segmentation. arXiv Preprint arXiv:1802.06955, 2018. https://doi.org/10.1109/NAECON.2018.8556686
3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 2018, 68(6), 394-424. https://doi.org/10.3322/caac.21492
4. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., Jemal, A. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 2024, 74(3), 229-263. https://doi.org/10.3322/caac.21834
5. Bruntha, P. M., Rose, D. J., Shruthi, A. T., Juliet, K. G., Kanimozhi, M. Application of Selective Region Growing Algorithm in Lung Nodule Segmentation. In Proceedings of the 2018 4th International Conference on Devices, Circuits and Systems (ICDCS), 2018, 319-322. IEEE. https://doi.org/10.1109/ICDCSyst.2018.8605155
6. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Zhou, Y., Yuille, A. L. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv Preprint arXiv:2102.04306, 2021. https://doi.

org/10.48550/arXiv.2102.04306

7. Fallahpoor, M., Nguyen, D., Montahaei, E., Rahmani, M., Behravan, H., Jones, B., Vinod, S. Segmentation of Liver and Liver Lesions Using Deep Learning. Physics in Medicine & Biology, 2024, 47, 611-619. https://doi.org/10.1007/s13246-024-01390-4. https://doi.org/10.1007/s13246-024-01390-4
8. Farag, A. A., El Munim, H. E., Graham, J. H., Farag, A. A. A Novel Approach for Lung Nodules Segmentation in Chest CT Using Level Sets. IEEE Transactions on Image Processing, 2013, 22(12), 5202-5213. https://doi.org/10.1109/TIP.2013.2282899
9. Guan, S., Khan, A. A., Sikdar, S., Chitnis, P. V. Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal. IEEE Journal of Biomedical and Health Informatics, 2020, 24(2), 568-576. https://doi.org/10.1109/JBHI.2019.2912935
10. Guo, S., Liu, X., Zhang, H., Lin, Q., Xu, L., Shi, C., Gao, Z., Guzzo, A., Fortino, G. Causal Knowledge Fusion for 3D Cross-Modality Cardiac Image Segmentation. Information Fusion, 2023, 99, 101864. https://doi.org/10.1016/j.inffus.2023.101864
11. Hou, J., Yan, C., Li, R., Huang, Q., Fan, X., Lin, F. Lung Nodule Segmentation Algorithm With SMR-UNet. IEEE Access, 2023, 11, 34319-34331. https://doi.org/10.1109/ACCESS.2023.3264789
12. Keshani, M., Azimifar, Z., Tajeripour, F., Boostani, R. Lung Nodule Segmentation and Recognition Using SVM Classifier and Active Contour Modeling: A Complete Intelligent System. Computers in Biology and Medicine, 2013, 43(4), 287-300. https://doi.org/10.1016/j.compbiomed.2012.12.004
13. Kostis, W. J., Reeves, A. P., Yankelevitz, D. F., Henschke, C. I. Three-Dimensional Segmentation and Growth-Rate Estimation of Small Pulmonary Nodules in Helical CT Images. IEEE Transactions on Medical Imaging, 2003, 22(10), 1259-1274. https://doi.org/10.1109/TMI.2003.817785
14. Krizhevsky, A., Sutskever, I., Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, 2017, 60(6), 84-90. https://doi.org/10.1145/3065386
15. Kuhnigk, J. M., Dicken, V., Zidowitz, S., Bornemann, L., Bakai, A., Wormanns, D., Krass, S., Peitgen, H.-O. Morphological Segmentation and Partial Volume Analysis for Volumetry of Solid Pulmonary Lesions in Thoracic CT Scans. IEEE Transactions on Medical Imaging, 2006, 25(4), 417-434. https://doi.org/10.1109/TMI.2006.871547
16. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, 9992-10002. IEEE. https://doi.org/10.1109/ICCV48922.2021.00986
17. Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., Xie, S. A ConvNet for the 2020s. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, 11966-11976. IEEE. https://doi.org/10.1109/CVPR52688.2022.01167
18. Mao, Q., Zhao, S. G., Gong, T., Xie, Y., Pan, J., Wang, Y. An Effective Hybrid Windowed Fourier Filtering and Fuzzy C-Means for Pulmonary Nodule Segmentation. Journal of Medical Imaging and Health Informatics, 2018, 8(1), 72-77. https://doi.org/10.1166/jmihi.2018.2235
19. Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., Rueckert, D. Attention U-Net: Learning Where to Look for the Pancreas. arXiv Preprint arXiv:1804.03999, 2018. https://doi.org/10.48550/arXiv.1804.03999
20. Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., Jiang, H. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. In Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023, 1-5. IEEE. https://doi.org/10.1109/ICASSP49357.2023.10096516
21. Qin, C., Wu, Y., Liao, W., Zeng, J., Liang, S., Zhang, X. Improved U-Net3+ With Stage Residual for Brain Tumor Segmentation. BMC Medical Imaging, 2022, 22(1), 14. https://doi.org/10.1186/s12880-022-00738-0
22. Rahman, M. M., Munir, M., Marculescu, R. EMCAD: Efficient Multi-Scale Convolutional Attention Decoding for Medical Image Segmentation. arXiv Preprint arXiv:2405.06880, 2024. https://doi.org/10.1109/CVPR52733.2024.01118
23. Ramamoorthy, H., Ramasundaram, M., Raja, S. P., Randive, K. An Efficient Classification of Multiclass Brain Tumor Image Using Hybrid Artificial Intelligence with Honey Bee Optimization and Probabilistic U-RSNet. International Journal of Image and Graphics, 2025, 25(01). https://doi.org/10.1142/S0219467824500591
24. Ramamoorthy, H., Ramasundaram, M., Raj, R. S. P., Randive, K. TransAttU-Net Deep Neural Network for

Brain Tumor Segmentation in Magnetic Resonance Imaging. IEEE Canadian Journal of Electrical and Computer Engineering, 2023, 46(4), 298-309. https://doi.org/10.1109/ICJECE.2023.3289609

25. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
26. Sangamithraa, P. B., Govindaraju, S. Lung Tumour Detection and Classification Using EK-Mean Clustering. In Proceedings of the 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, 2201-2206. IEEE. https://doi.org/10.1109/WiSPNET.2016.7566533
27. Schlemper, J., Oktay, O., Chen, L., Matthew, J., Knight, C., Kainz, B., Glocker, B., Rueckert, D. Attention-Gated Networks for Improving Ultrasound Scan Plane Detection. arXiv Preprint arXiv:1804.05338, 2018. https://doi.org/10.48550/arXiv.1804.05338
28. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 2021, 71(3), 209-249. https://doi.org/10.3322/caac.21660
29. Yang, T., Cheng, J., Zhu, C. A Segmentation of Pulmonary Nodules Based on Improved Fuzzy C-Means Clustering Algorithm. MATEC Web of Conferences, 2018, 232, 03011. https://doi.org/10.1051/matecconf/201823203011
30. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. IEEE Transactions on Medical Imaging, 2020, 39(6), 1856-1867. https://doi.org/10.1109/TMI.2019.2959609