IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 November 2024, accepted 23 November 2024, date of publication 26 November 2024,
date of current version 5 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3506773

RESEARCH ARTICLE

MDE-Net: Multi-Layer Depth Extraction Network With Attention Mechanism for Medical Image Segmentation

XIAOKANG DING, LING DONG, YINGYU JI, AND KE'ER QIAN
College of Mechanical Engineering, Quzhou University, Quzhou 324000, China

Corresponding author: Ling Dong (dongling@qzc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62102227, in part by Zhejiang Basic Public Welfare Research Project under Grant LZY24E050001, Grant LZY24E060001, and Grant ZCLTGS24E0601; and in part by the Science and Technology Major Projects of Quzhou under Grant 2022K56, Grant 2023K221, and Grant 2023K211.

ABSTRACT Accurate segmentation of organ and pathological tissue images is of great significance to the diagnosis and treatment of various diseases. However, it still faces great challenges due to the inherent complexity, diversity, noise and occlusion of medical image data. To solve these problems, based on the U-Net framework, we propose a medical image segmentation algorithm named MDE-Net, which combines multi-layer deep feature extraction module and attention mechanism. Firstly, in the encoding part, we introduce the hybrid convolutional feature extraction (HCFE) module as a replacement for traditional convolutional blocks to allow for a more robust extraction of features at multiple scales and help expand the receptive field. Subsequently, we design the multi-layer pooling and channel-spatial squeeze & excitation (MPcsSE) module, which extracts more image context information by multi-layer pooling connection of the coding part and introducing csSE module in the middle connection part. Finally, in the decoder part, we design the SE-MultiResConv that combines multi-scale residual convolution with SE attention mechanism to improve segmentation accuracy and prevent the loss of detail information during up-sampling. In extensive experiments, we conducted detailed tests on two publicly available medical image datasets to rigorously evaluate the performance of our proposed MDE-Net. For the ISIC-2018 dataset, MDE-Net achieved remarkable metrics with Accuracy of 91.59%, Matthews correlation coefficient (Mcc) of 81.78%, Dice of 86.63%, and Jaccard of 76.98%. Similarly, on the COVID-19 dataset, MDE-Net exhibited outstanding performance, achieving Accuracy of 95.53%, Mcc of 79.92%, Dice of 83.43%, and Jaccard of 70.93%. The excellent performance of MDE-Net on these datasets proves its effectiveness and generalization in medical image segmentation tasks. By delivering precise and dependable segmentation outputs, MDE-Net demonstrates a transformative potential for the diagnosis and treatment of diverse medical conditions. MDE-Net's contribution can significantly streamline diagnostic processes, minimize human error, and optimize resource allocation in clinical settings, making it a valuable tool in advancing healthcare.

INDEX TERMS Medical image segmentation, multi-layer depth extraction, attention mechanism, squeeze-and-excitation, multi-layer pooling, multi-scale residual convolution, U-Net.

I. INTRODUCTION

Since the birth of digital medical imaging, the application of image processing techniques in the field of medical image analysis has attracted much attention, and it has become an important tool for doctors to analyze pathological tissues. Traditionally, the task of identifying and delineating pathological tissues has relied entirely on the expertise of clinicians. This manual approach presents several challenges, including the complexity of the task, the

The associate editor coordinating the review of this manuscript and approving it for publication was Tai Fei.

extensive time required for processing, and the variability in judgments among different observers. To address these multifaceted challenges, interdisciplinary researchers have been diligently working on developing automated diagnostic systems leveraging advanced technologies such as machine learning and deep learning. By integrating sophisticated image processing algorithms and powerful computational techniques, these automated diagnostic systems not only speed up the diagnostic process, but also provide clinicians with precise and detailed insights into pathological conditions and they significantly improve the efficiency and accuracy of pathological tissue analysis.

In the early days of medical image segmentation, traditional image processing methods such as edge detection, template matching techniques, statistical shape models, and active contours were primarily used. However, due to the complexity of medical images and the difficulty of representing their features, these methods still face significant limitations. With the improvement of computer performance and digital image processing technology, the scope and depth of medical image segmentation research are expanding, especially the emergence of deep learning algorithms [1], [2], [3]. These algorithms brought remarkable advancements with their superior feature extraction capabilities, highly flexible adaptability, and robust generalization ability. Currently, the most widely discussed are the fully convolutional networks (FCN) [4] and U-Net [5], both of which can achieve shallow and deep feature extraction with high accuracy and efficiency. However, due to the complexity and diversity of image segmentation data, these methods often struggle to extract comprehensive global information from feature images, which is crucial for the accurate analysis of pathological tissues.

In recent years, more and more structure optimization methods have been proposed, including multi-scale feature extraction [6], [7], attention mechanism [8], [9], and residual connection [10], [11]. Among them, Selvaraj et al. [12], [13] proposed a CRPU-Net to speed up colonoscopy, reduce unnecessary biopsies, improve patient care, and optimize medical resource allocation. Singh et al. [14] proposed a method of de-noising medical images to achieve accurate detection and classification, and studied how to use high performance computing to process massive image data to classify the effects of fluid accumulation and cartilage erosion in MRI images of knee joints with osteoarthritis. Khan and Singh [15] implemented semantic segmentation by modifying the feature space in the basic U-Net architecture to achieve the geometric features required for nonlinear road extraction, while striving to maintain appropriate edges and boundaries. Zhang et al. [16] suggested a medical image segmentation method based on convolutional attention blocks, which reduces the interference of invalid targets and achieves more comprehensive and effective feature extraction. Yin et al. [17] introduced the AMSUnet framework, a novel approach designed to enhance the segmentation of various medical targets. This framework employs a combination of subscale and multi-scale convolution reconstruction subsampling encoders, which work synergistically to improve segmentation accuracy and robustness. Zhang et al. [18] utilized the strengths of ConvNeXt and U-Net to propose an innovative approach named BCU-Net, designed to effectively handle a wide range of medical images with varying resolutions. Li et al. [19] fully capitalized on the strengths of CNNs and Transformer to introduce a dual-codec structure of X-Net. This synergistic integration allows X-Net to excel in capturing both local details and long-range dependencies, which are essential for accurately segmenting complex medical images. Zheng et al. [20] introduced a novel network called CASF-Net, which adeptly integrates both coarse-grained and fine-grained feature representations. This innovative architecture leverages cross-attention and cross-scale fusion mechanisms to effectively capture and fuse features from different levels of abstraction. Xie et al. [21] developed an advanced architecture incorporating a two-stream pyramid module alongside a context-aware encoder-decoder module. The design goal is to enhance the learning of local detail features across multiple scales, which is essential for effectively distinguishing between intricate medical images and their complex backgrounds. Despite the continuous development of various innovative algorithms, how to accurately segment lesion areas of medical images remains a major challenge for researchers. For example, the ISIC-18 dataset covers a variety of skin lesion types, but each lesion may be significantly different in shape, color, texture, and the boundary of some skin lesions has a transition zone and a mixed area of other normal skin, which increases the difficulty of segmentation. Similarly, in the context of the COVID-19 datasets, the infection areas are scattered and contain many disjointed boundaries, which often makes it difficult for traditional segmentation algorithms to produce clear and distinct boundaries, resulting in fuzzy segmentation results. Given these multifaceted challenges, it becomes critical that researchers develop new techniques to address the diversity and complexity of image segmentation in modern medicine.

Inspired by the aforementioned methodologies, we proposed a multi-layer deep feature extraction network with attention mechanism for medical image segmentation, termed MDE-Net. Specifically, in the encoding stage, we have replaced traditional convolutional blocks with a hybrid convolutional feature extraction module. At the juncture between the encoder and decoder, we have introduced a csSE attention mechanism that combines with multi-layer pooling. In the decoding phase, we have incorporated a SE attention mechanism alongside a multi-scale residual convolution module. Our extensive experiments on the ISIC-2018 dataset and COVID-19 dataset demonstrated that the proposed MDE-Net had significant segmentation performance. The principal contributions of our work are as follows:

(1) By capturing fine-grained patterns alongside larger contextual cues, the HCFE blocks ensure a more comprehensive and robust feature extraction process. This design collaboratively combines standard and extended convolution techniques to improve the feature representation capability of the network, especially in complex segmentation tasks.

(2) The MPcsSE block uses a refined feature processing strategy to effectively suppress noise and redundant information in images through multi-level and multi-scale pooling operations. This novel integration effectively extracts and emphasizes contextual information, and it ensures that the network maintains a high level of accuracy in identifying and segmentation key features at different scales and spatial locations.

(3) The SE-MultiResConv module can capture feature information of different scales by adding multi-scale residual convolution, and effectively alleviate the problem of gradient disappearance through residual connection. This combination not only improves the segmentation accuracy, but also prevents the loss of detail information in the up-sampling process.

The overall composition of this paper is as follows: In the second section, we present a detailed explanation of the proposed MDE-Net method, including its architecture and core components. In the third section, we describe the performance of MDE-Net compared to the most advanced methods. Finally, the last section briefly summarizes the results of the study.

II. METHODS

In this section, we propose a method that combines multi-scale feature extraction with attention mechanisms to segment medical images. Our method is composed of three parts: encoder module, intermediate connection module and decoder module. In the following subsection, we will explore these modules in depth.

A. FRAMEWORK OF MDE-Net

s In the field of medical image segmentation, the encoder-decoder structure in the U-Net architecture is particularly effective in capturing semantic information. Its skip connection is able to combine the low-level features of the encoder with the high-level features of the decoder. However, as a relatively simple network, U-Net may struggle to handle complex scenes or capture finer semantic details. To address these limitations, we have developed a new module based on the U-Net architecture specifically designed to optimize segmentation performance for more challenging tasks and finer image detail extraction, as shown in Fig. 1. Specifically, to ensure the efficient extraction of fundamental features, our approach employs hybrid convolutional feature extraction modules within the first four layers of the encoder. Then, through a series of down-sampling operations with step size of 2, the input features are divided into four different convolution blocks layer by layer. For the input image $I \in R^{3\times H\times W}$, each convolutional block of extracted HCFE features is represented as $R^{c_t\times h_t\times w_t}$, where the channel number c_t values followed by $\{16, 32, 64, 128\}$, $h_t = H/2^t$, and $w_t = W/2^t$. In the final layer, we retain the original 3×3 convolutional block. To efficiently capture a wider context and ensure reliable extraction of image features, we developed the MPcsSE module. The module is designed to integrate feature information extracted from the first to fourth layers of the encoder, and it leverages both low-level and high-level features to create a comprehensive representation of the input image. In each step of the expansion path, the process involves up-sampling the feature map, followed by a connection to the corresponding cropped feature map from the shrink path. Unlike the standard approach, where a 3×3 convolution is applied, we substitute it with an SE-MultiResConv module. This modification serves to reduce the number of feature channels by half while simultaneously enhancing feature representation. Finally, we output the final feature image through a 1×1 convolution operation. In the subsequent sections, we will delve deeper into each of these modules, illustrating how they integrate and contribute to the overall network.

B. HYBRID CONVOLUTIONAL FEATURE EXTRACTION BLOCK

In our proposed MDE-Net, we designed a hybrid convolutional feature extraction block that enhances feature extraction by integrating multiple convolutional operations. As illustrated in Fig. 2, the input feature map undergoes several parallel processing paths to capture diverse and comprehensive features. Firstly, the input feature is divided into three branches, with the first two branches working in parallel and independently processing the input feature maps. Each of the two branches first carries out convolution operation through a standard 3×3 convolution kernel, extracting local details of input features from different perspectives, and providing a rich information basis for the subsequent feature fusion. After the local feature extraction is completed, the feature maps of these two branches are combined. The combined feature map is processed through a 1×1 convolution layer for dimensionality reduction while preserving the most important essential features. The third branch adopts dilated convolution, which increases the receptive field by inserting "holes" between the elements of the standard convolution kernel without the need for additional parameters or computational complexity. This feature enables the network to capture a wider range of contextual information without losing image resolution. Finally, the feature maps of different scales from the three branches are combined to enable the network to simultaneously obtain local details and global contextual information, thereby enhancing the characterization ability of the network. In short, feature maps of different scales complement each other during the fusion process, enabling the network to more accurately understand complex medical image structures and improve the performance of tasks such as segmentation and detection.

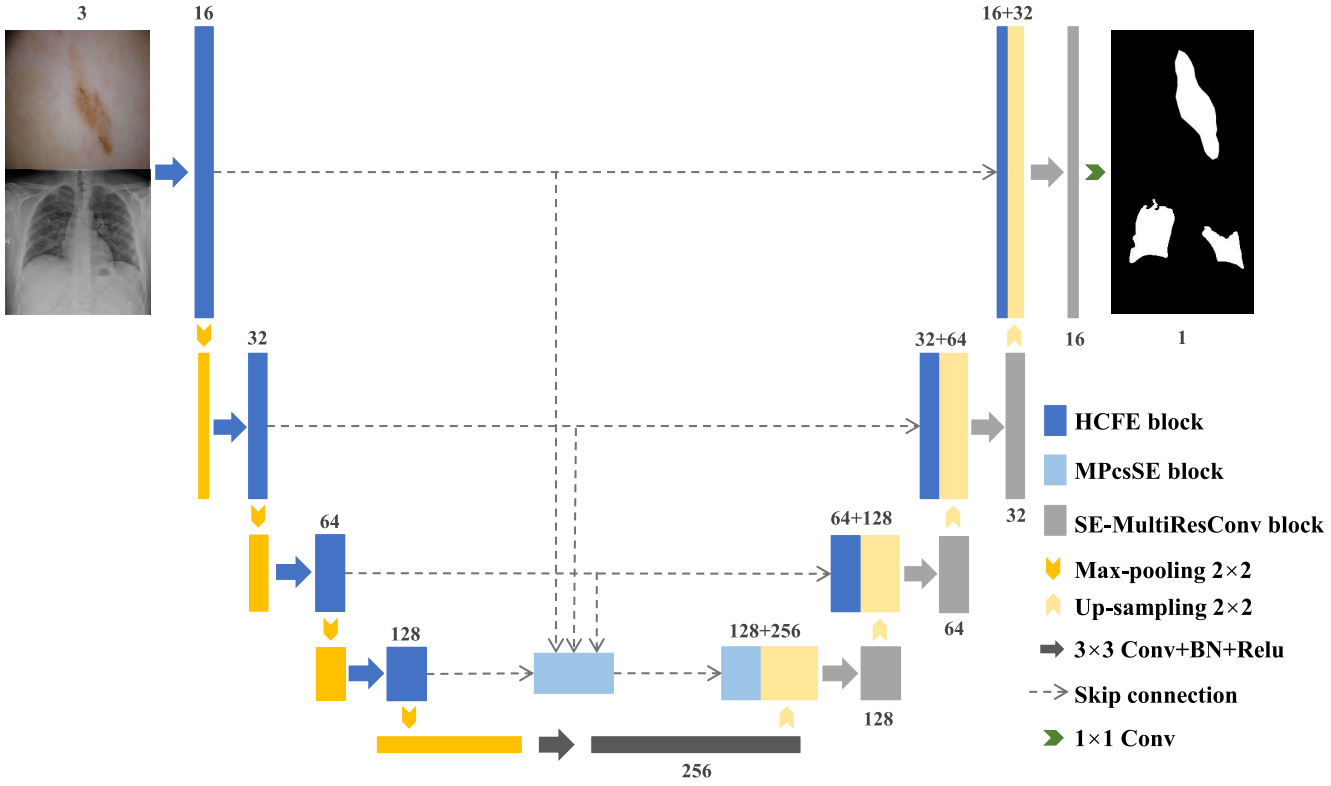

FIGURE 1. Network architecture of MDE-Net.

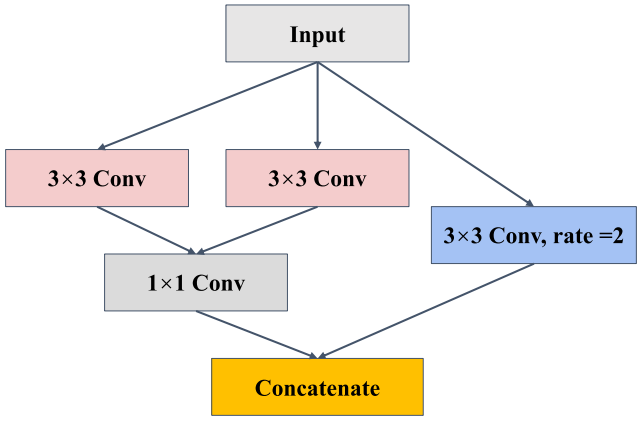

FIGURE 2. Structure of hybrid convolutional feature extraction block.

C. MULTI-LAYER POOLING AND CHANNEL-SPATIAL SQUEEZE & EXCITATION

In order to more effectively address the challenge of scarce and diverse lesion area samples in medical image analysis, we designed a multi-layer pooling module, as illustrated in Fig. 3. This module aims to effectively reduce image noise and redundant information through refined feature processing strategies, while enhancing the ability to extract global features. In MPcsSE block, we carefully planned four parallel processing paths, each of which performs pooling operations at different scales for feature images at different levels. First of all, the first path applies the max-pooling directly to the first convolution feature image output by the encoder, rapidly reducing the resolution of the feature map to 1/8 of the original size. The network can quickly capture the significant features in the image and initially reduce the data dimension, reducing the burden for subsequent processing. The second path applies max-pooling again to the feature image after the second convolution, reducing the resolution to 1/4 of the original size to preserve more details. The third path continues to perform max-pooling on the feature image after the third convolution, but the pooling step size is reduced, and the resolution is only reduced to 1/2 of the original size. The feature map of this path achieves a good balance between global and local information, which helps the model to understand the image content more comprehensively. The fourth path keeps the feature image output from the fourth convolution unchanged without any additional pooling operations. This path preserves the highest resolution feature map and is essential for capturing subtle lesion changes. Upon completing the multi-layer pooling operations across these four paths, we integrate the resulting feature maps. To further enhance the model's adaptive capacity and feature filtering capability, we introduce the csSE attention mechanism. This attention mechanism recalibrates the fused feature map, selectively enhancing useful features while suppressing irrelevant ones. In summary, the MPcsSE

module substantially enhances the model's ability to acquire comprehensive contextual information, thereby providing robust support for the accurate analysis and diagnosis of medical images. This innovative design not only boosts the overall segmentation performance but also ensures that the network remains efficient and resilient.

The csSE attention mechanism, as shown in Fig. 4, is to deform the SE attention module [22], [23]. The csSE mechanism achieves this by decomposing the SE module into two distinct components: the channel-wise SE (cSE) module and the spatial-wise SE (sSE) module. Each component processes the input feature maps independently to produce calibrated feature maps, which are subsequently combined to form the final enhanced feature map. Specifically, the cSE module first passes the input feature map through the global averaging pooling layer, changing the dimension from [C, H, W] to [C, 1, 1], where C is the number of channels, H and W are the height and width of the feature map respectively. Following this, the reduced feature map undergoes a transformation through two 1×1 convolutional layers. The first convolutional layer reduces the number of channels to 1/r of the original size, where r is a reduction ratio parameter that controls the extent of dimensionality reduction. The second convolutional layer then restores the number of channels back to the original size. The result of these convolutions is then passed through a sigmoid activation function, normalizing the values to produce a channel-wise weight vector. This weight vector is subsequently multiplied with the original feature map in a channel-wise manner, resulting in a feature map that has been calibrated according to the importance of each channel. In parallel, the sSE module processes the input feature map by applying a 1×1 convolutional layer directly to it. This layer sets the number of output channels to one and uses a convolution kernel of size 1×1, effectively compressing the channel dimension while retaining the spatial structure of the feature map. The resulting spatial dimension weight map is then normalized using a sigmoid activation function, producing a spatial-wise weight map. In addition, this weight map is multiplied element-wise with the original feature map, calibrating the spatial information by emphasizing the most relevant spatial features. The final step in the csSE attention mechanism involves combining the outputs of the cSE and sSE modules. This combined approach allows the csSE module to leverage both spatial and channel-wise information, providing a more comprehensive and nuanced representation of the input feature map. The specific calculation formula of cSE and sSE are as follows:

$$z_k = \frac{1}{H \times W} \sum_{\mathrm{i}}^{H} \sum_{j}^{W} u_k (i, j) \tag{1}$$

$$\hat{U}_{cSE} = F_{cSE}(U) = \left[\sigma\left(\hat{z}_1\right) u_1, \sigma\left(\hat{z}_2\right) u_2, \cdots, \sigma\left(\hat{z}_c\right) u_c\right] \tag{2}$$

$$\hat{U}_{sSE} = F_{sSE}(U)$$

$$= \left[\sigma\left(q_{1,1}\right) u^{1,1}, \cdots, \sigma\left(q_{i,j}\right) u^{i,j}, \cdots \sigma\left(q_{H,W}\right) u^{H,W}\right] \tag{3}$$

D. MULTI-SCALE RESIDUAL CONVOLUTION WITH SE ATTENTION MECHANISM

In the process of deepening and optimizing the U-Net architecture, the decoder part plays a key role in decoding the highly compressed encoder feature vector and reconstructing it into the original image size and content. However, the traditional decoding strategy mainly relies on simple deconvolution and upsampling operations, which often result in the loss of image details while increasing the spatial resolution of feature maps, leading to problems such as image blurring, unclear edges, and even information distortion. To overcome this limitation, we propose a SE-MultiResConv module, which combines multi-scale residual convolution with SE attention mechanism, as shown in Fig. 5, aiming to significantly improve the retention and enhancement of feature information during decoding. The innovation of this module comes from the dual pursuit of efficiency and effectiveness of feature extraction. By processing two feature extraction paths in parallel, the module realizes comprehensive and precise analysis of input features. The first path focuses on using SE attention mechanism to improve the feature representation ability of the model. The SE attention mechanism, as shown in Fig. 6, automatically learns the importance of each feature channel through two key operations: squeeze and excitation. Based on these two operations, the feature channel is re-labeled to enhance the important features and suppress the unimportant ones. The second path focuses on capturing multilevel features in images through the combination of multi-scale convolution and residual connection. In this path, the input images are first processed by 1×1, 3×3 and 5×5 convolutions respectively to capture local features of different scales. This multi-scale convolution strategy not only enriches the diversity of features, but also helps models better understand and represent complex structures in images. Then, the output of different convolutional layers is fused by residual connection, which not only preserves the details of shallow features, but also takes advantage of the abstract representation ability of deep features. The residual connection also effectively alleviates the problem of gradient disappearance or explosion in deep network training, accelerates the training process, and improves the scalability of the network. Finally, SE-MultiResConv block fuses the output of the two paths, and achieves the complementary and enhanced feature information through the superposition operation of feature graphs. This fusion process makes full use of the advantages of SE attention mechanism in feature re-calibration, and combines the richness and robustness of multi-scale convolution residual module in feature extraction. Therefore, it can extract multi-layer detailed features from complex data more effectively, and provide more abundant and accurate feature representation for subsequent image reconstruction tasks.

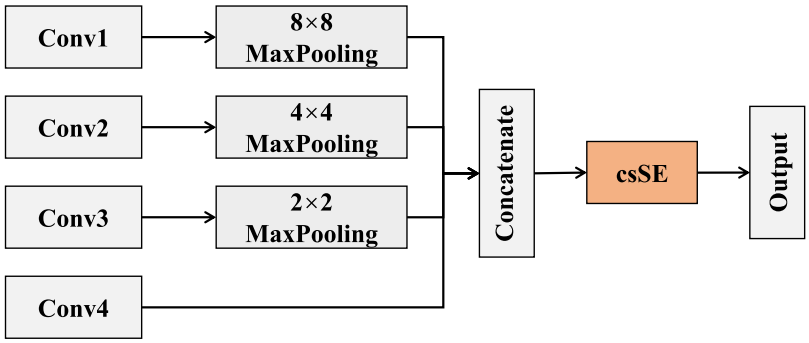

FIGURE 3. Structure of multi-layer pooling and channel-spatial squeeze & excitation block.

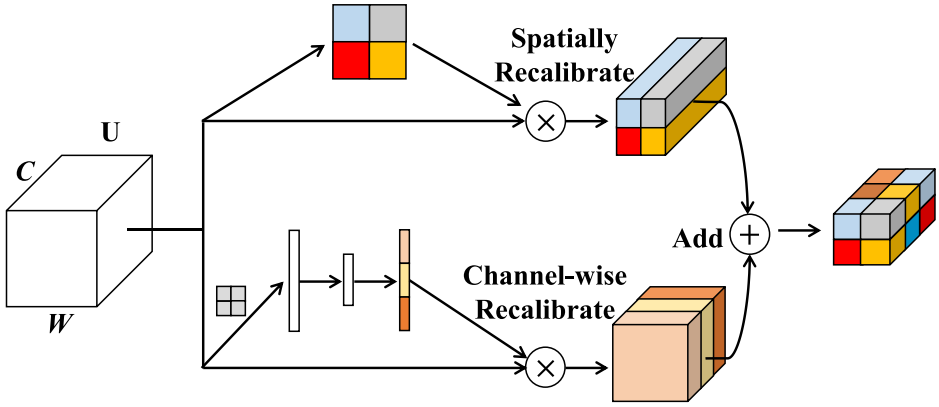

FIGURE 4. Structure of channel-spatial squeeze & excitation module.

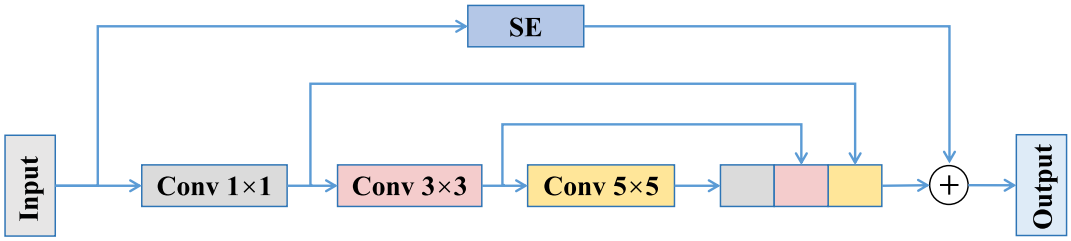

FIGURE 5. Structure of multi-scale residual convolution with SE attention mechanism.

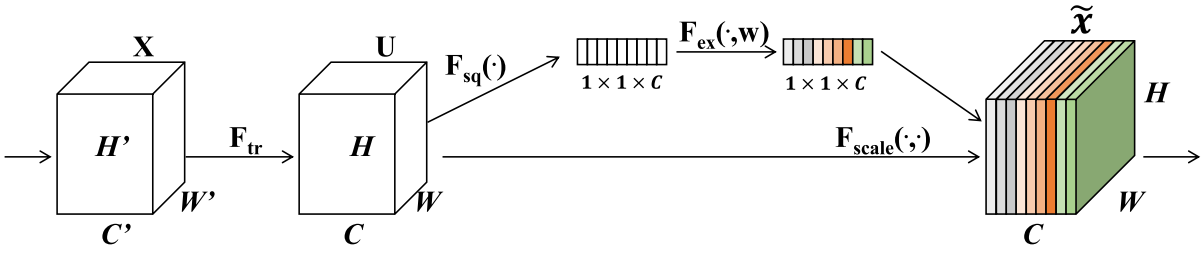

FIGURE 6. Structure of SE attention module.

E. LOSS FUNCTION

Due to the limitations of small datasets, coupled with the complexity of lesion size and the uneven distribution between foreground and background in the medical images, the problem of category imbalance becomes particularly prominent. Compared with the traditional cross entropy loss

function, Dice loss [24], [25] is widely used because of its unique advantages that it can handle the class imbalance problem well and still produce accurate segmentation effect in the case of limited labeled data. Therefore, the Dice loss function is chosen in this paper, and its specific calculation method is described as follows:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \hat{y}_i^2} \tag{4}$$

III. RESULTS

A. DATASET

In this section, we evaluate the performance of MDE-Net on both the ISIC-18 dataset and the COVID-19 dataset. Recognizing the importance of standardization in the data preprocessing phase, we put each image through a rigorous adjustment procedure, with a size of 256 × 256 pixels. The details of these data sets are as follows.

1) ISIC-2018 DATASET

ISIC-2018 is currently the world's largest skin lesion image dataset with professionally annotated information attached to digital skin lesion images. The dataset contains 3,694 images, of which 2,594 for training, 100 for validation, and 1,000 for rigorous testing to assess the generalization and feature extraction capabilities of our model. The dataset can be downloaded at: https://challenge.isic-archive.com/data/#2019. Fig. 7(a-b) provides some original images of the ISIC-18 dataset and their corresponding annotated images.

2) COVID-19 DATASET

The medical image acquisition process is complex and requires high standards of accuracy and trust in the marks. However, due to the scarcity of training samples, the challenge of over-fitting is often faced in the training process, which weakens the prediction accuracy of the algorithm. In order to mitigate the risk of training data scarcity and overfitting, we used the COVID-19 dataset in addition to the ISIC-2018 dataset, so that the training data can be diversified on the basis of ensuring semantic integrity and lossless. In order to more accurately understand the generalization ability and robustness of the model on different pathological features and data sets. The COVID-19 dataset contains 2,913 COVID-19 images, 1,864 of which were used for training, 466 for validation, and 583 images were processed for testing. The dataset can be obtained from: https://www.kaggle.com/datasets/anasmohammedtahir/covidqu. Fig. 7(c-d) provides some original images of the COVID-19 dataset and their corresponding labeled images.

B. THE TRAINING AND VERIFICATION OF MDE-Net

The network architecture is deployed on a 64-bit Windows operating system and relies on the Python 3.7.0 library for development. To speed up the calculation process, the system is equipped with an NVIDIA Quadro RTX 6000 high-end graphics processing unit (GPU) with 24GB of extended memory capacity. During the training process, we adopted the following key hyper-parameter configuration: the number of samples per batch is 16, the entire training process will be repeated for 200 iterations. The initial value of the learning rate is set to 0.001 to optimize the convergence rate and performance of the model. The selection of this value is the result of in-depth understanding of the model training dynamics and many experiments, and ensures that it can converge smoothly to the optimal solution in the late training period. In the selection of optimizer, we deeply analyzed the mainstream optimization algorithms in the current deep learning field, and finally decided to use Adam optimizer [26], [27]. With its unique advantages, Adam optimizer can adjust the learning rate adaptively in the process of parameter updating, which significantly speeds up the training speed and improves the convergence efficiency of the model. To mitigate the negative effects of model overfitting, we implemented a well-designed early termination mechanism. When the performance on the validation set is no longer increasing, it interrupts the training process to prevent the model from falling into the overfitting region, thus ensuring good generalization ability of the model. Fig. 8 shows the accuracy and loss curves of the proposed MDE-Net on ISIC-2018 dataset and COVID-19 dataset, where blue represents the training curve and orange represents the validation curve. It can be seen from the graphical data that MDE-Net has outstanding performance in the whole training and verification stage, and the convergence speed is also extremely efficient, perfectly avoiding the influence of overfitting or under-fitting.

C. EVALUATION METRICS

In this paper, four key indicators, Dice [28], [29], Mcc [30], [31], Accuracy [32], [33] and Jaccard [34], [35], are used to evaluate the performance of each model. Specific calculations are as follows:

$$Dice = \frac{2TP}{2TP + FN + FP} \tag{5}$$

$$Mcc = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FN)\,(TP+FP)\,(TN+FP)\,(TN+FN)}} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Jaccard = \frac{TP}{TP + FN + FP} \tag{8}$$

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

D. COMPARISON WITH OTHER METHODS

To emphasize the superiority of our approach, the proposed MDE-Net was conducted in-depth comparative analysis with

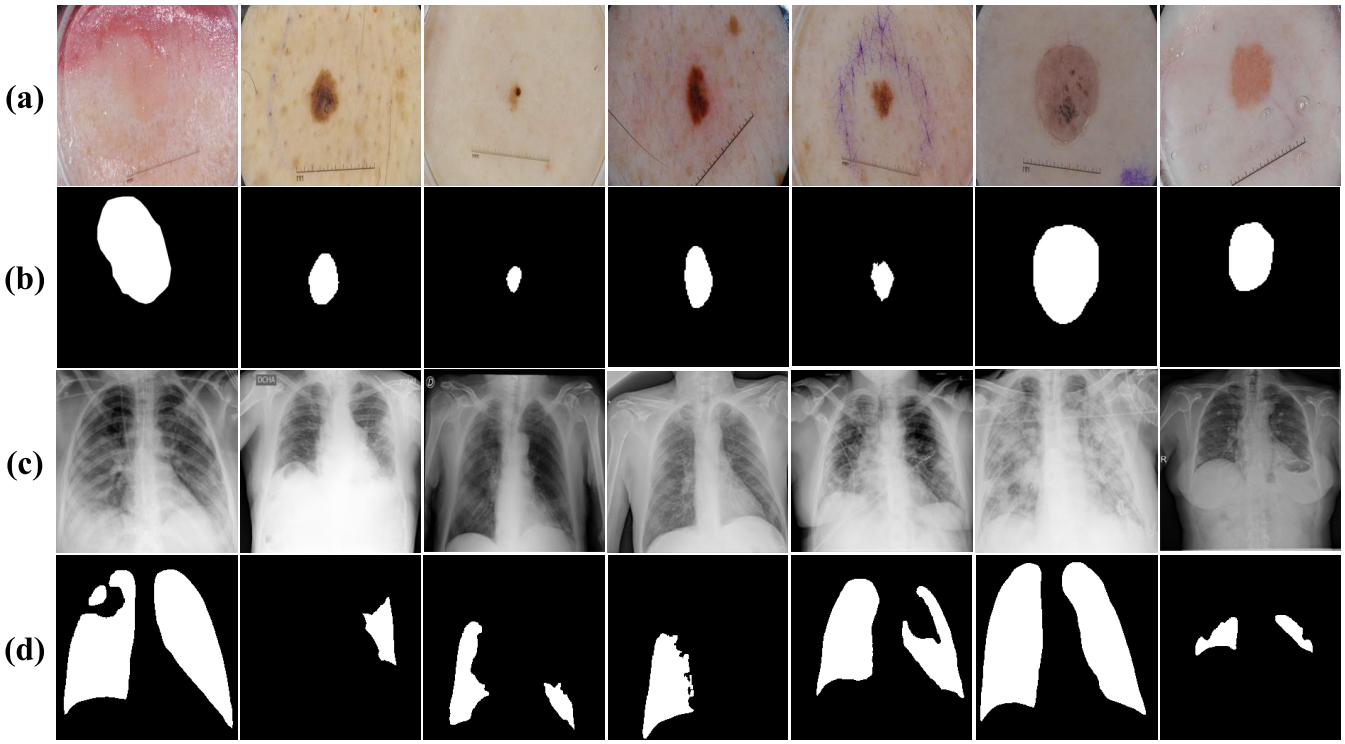

FIGURE 7. Some original images and their corresponding labels. (a-b) images from ISIC-18 dataset and their corresponding annotated. (c-d) images from COVID-19 dataset and their corresponding labeled.

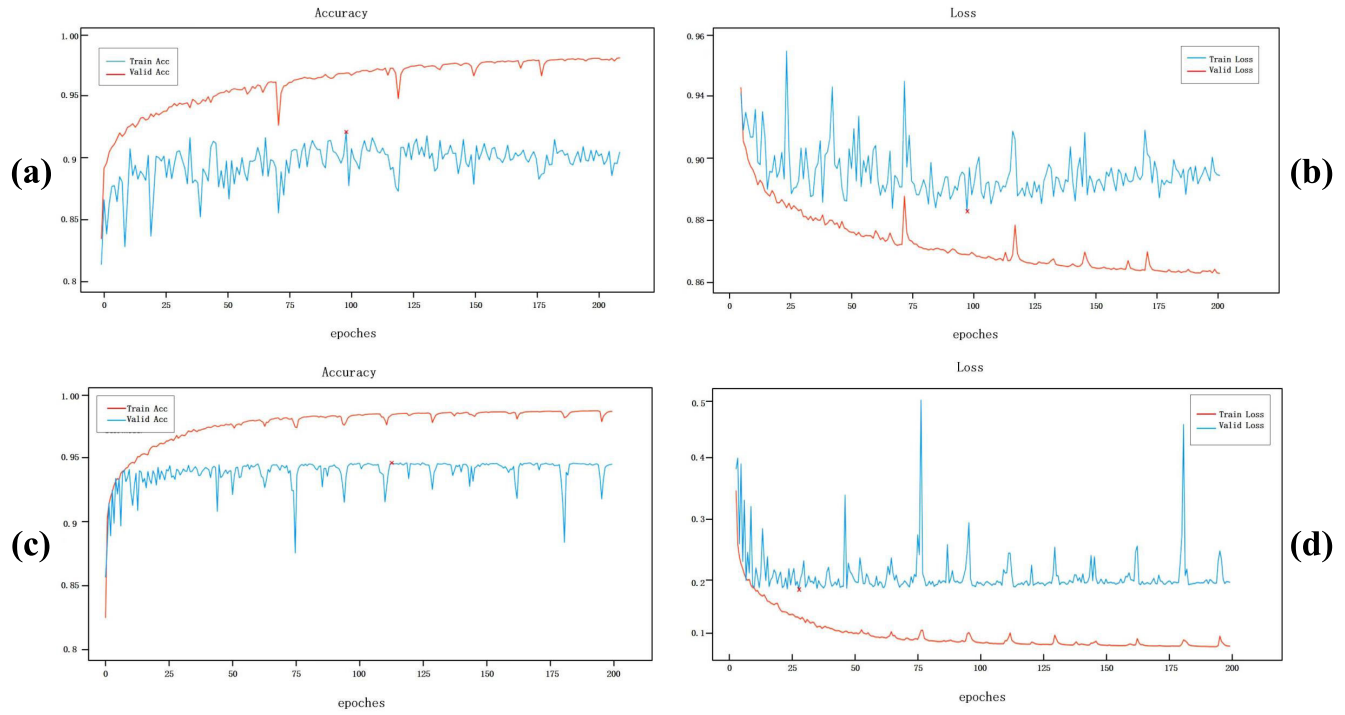

FIGURE 8. Accuracy and loss curve of the proposed MDE-Net on ISIC-2018 dataset and COVID-19 dataset. (a-b) accuracy and loss curves on ISIC-18 dataset. (c-d) accuracy and loss curves on COVID-19 dataset.

multiple frameworks such as U-Net, AttUNet, R2UNet, HRNet, CLNet, Connected_UNet, ODsegmatiton, SEUNet, EENet, and ASFNet on the ISIC-2018 dataset. These models represent the vanguard of image segmentation techniques.

TABLE 1. Comparison experiment on the ISIC-18 dataset.

Method	Dice(%)	Mcc(%)	Accuracy(%)	Jaccard(%)
U-Net [5]	82.77	76.71	90.67	71.24
AttUNet [36]	84.25	78.41	90.56	73.38
R2UNet [37]	83.10	77.13	90.31	71.68
HRNet [38]	83.95	77.99	90.89	72.87
CLNet [39]	81.38	74.80	89.97	69.25
Connected_UNet [40]	85.43	80.00	90.63	75.21
ODsegmatiton [41]	81.82	75.61	91.30	69.94
SEUNet [42]	82.28	75.86	90.79	70.50
EENet [43]	84.55	78.84	90.78	73.68
ASFNet [44]	86.21	81.26	91.28	76.47
SCTV-UNet [45]	83.32	77.16	90.73	72.03
SSA-UNet [46]	85.12	79.46	91.56	74.67
RMAU-Net [47]	85.04	79.53	90.95	74.53
MDE-Net	86.63	81.78	91.59	76.98

After evaluation, its performance is detailed in Table. 1. It is necessary to note that although models such as U-Net, CLNet, ODsegmentation, SEUNet, SCTV-UNet and RMAU-Net have shown impressive performance in their respective fields, they appear inadequate and perform relatively poorly when facing the rich details, varied textures, and complex pathological features in skin cancer images. This is mainly due to the limitations of these models in feature extraction and analysis, which make it difficult to fully capture and understand subtle changes in images. Using channel attention and self-attention mechanisms, SSA-UNet can automatically weight different feature mappings, and it shows strong performance, particularly in accuracy metrics. Instead, ASFNet deepens the spatial upsampling block and proposes a multi-scale convolutional block that improves its adaptability and ability to recognize complex details. However, even so, it still fails to achieve optimal performance on all evaluation indicators. Among these networks, our MDE-Net has the most outstanding performance, with Dice, Mcc, Accuracy and Jaccard scoring 86.63%, 81.78%, 91.59% and 76.98% respectively. Compared with U-Net, these indicators have improved significantly, with profit margins of 3.86%, 5.07%, 0.92% and 5.47%, respectively. The experimental results show that the HCFE block, MPcsSE block and SE-MultiResConv block proposed in this paper are effective, which can realize the extraction of global network feature information and improve the segmentation accuracy.

To visually evaluate and compare segmentation performance, the visualized results of our model and various typical models are shown in Fig. 9. As can be seen from the figure, U-Net showed obvious inaccuracy in the segmentation of skin cancer lesion images, resulting in discontinuous segmentation (the third row of Fig. 9). CLNet, R2UNet, ODsegmatition, and SEUNet were inspired by the ''pruning'' module, residual connection, and attention mechanism. These models achieved similar performance to U-Net (row 7, 5, 9, and 10 of Fig. 9), but limited by insufficient sensitivity field and high segmentation complexity for small images. There are still gaps. To address these challenges, AttUNet is seen as a variant of U-Net, with attention gates and skip connections, to improve segmentation accuracy and efficiency, as shown in the fourth row of Fig. 9. More recently, HRNet has further guided network segmentation by introducing multi-branch parallelism and multi-scale fusion in U-Net (row 6 of Fig. 9). To further precise segmentation of skin cancer lesions, EENet incorporates multi-layered residual network connections to widen the acceptance field and mitigate gradient disappearance. However, as can be seen from the eleventh row of Fig. 9, the segmentation results of EENet still have discontinuity and wrong segmentation, which is caused by incomplete global information extraction during the encoding and decoding connection. In the eighth row of Fig. 9, Connected_UNet improves the accuracy of fine segmentation by leveraging tighter connections. In the twelfth row of Fig. 9, ASFNet achieves the second best result by combining adaptive structure and feature fusion techniques. SCTV-UNet, SSA-UNet, and RMAU-Net demonstrate noticeable challenges in segmentation accuracy, as evidenced by a significant number of missegmentations in their outputs (row 13, 14, and 15 of Fig. 9). In contrast, our approach enhances the coverage of receptive fields and integrates multi-scale feature maps effectively, which is mainly due to our proposed HCFE block and MPcsSE block. In this way, both local and global information can be obtained through multi-scale fusion, and the adaptive ability of the model can be improved. In addition, SE-MultiResConv blocks can accurately segment fine features and extract multiple layers of detailed features. The resulting segmentation results are more accurate than the other methods, as shown in the last row of Fig. 9.

In addition, we further demonstrated the performance of our proposed MDE-Net method by conducting a large number of experiments on COVID-19 dataset. Qualitative visualization is shown in Fig. 10, and quantitative analysis is listed in Table. 2. It can be easily seen that COVID-19 datasets have more complex backgrounds, which brings higher challenges to image segmentation, making it difficult to avoid varying degrees of errors in the segmentation process of various models. As a result, the Dice, Mcc, Accuracy and Jaccard values were significantly reduced. Despite these challenges, our proposed approach outperforms all competing models on all evaluation metrics, with an impressive Dice score of 83.43%, Mcc score of 79.92%, Accuracy score of 95.53%, and Jaccard score of 70.93%. This superior performance underscores the robustness and generalization ability of the MDE-Net method and highlights its accuracy in efficiently handling the complexity of the COVID-19 dataset. Therefore, in COVID-19 image segmentation, it shows advantages in accurate segmentation of small features and extraction of global feature information.

E. COMPUTATIONAL EFFICIENCY

In Table. 3, we conducted a comprehensive analysis of the relevant parameters and computational efficiency of each segmentation method on the ISIC-18 dataset. Notably, models such as U-Net, ASFNet, SCTV-UNet, SSA-UNet, RMAU-Net, and Connected_Net show clear advantages in requiring fewer parameters. However, this reduction in

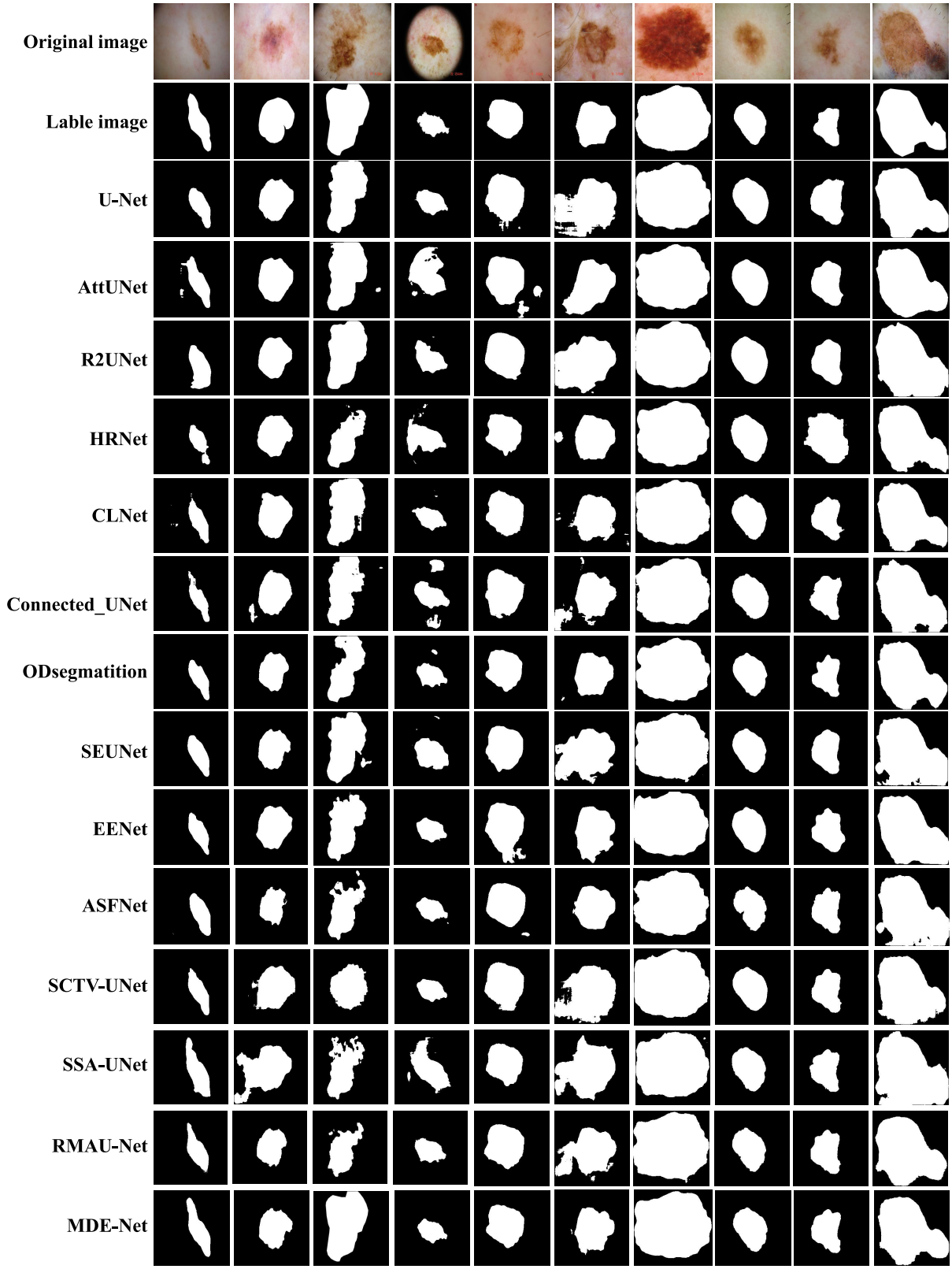

FIGURE 9. Visual segmentation results of different methods on the ISIC-18 dataset. First and second rows: the original images and their corresponding ground truth. The third to last rows are the result of U-Net, AttUNet, R2UNet, HRNet, CLNet, Connected_UNet, ODsegmatition, SEUNet, EENet, ASFNet, SCTV-UNet, SSA-UNet, RMAU-Net and MDE-Net.

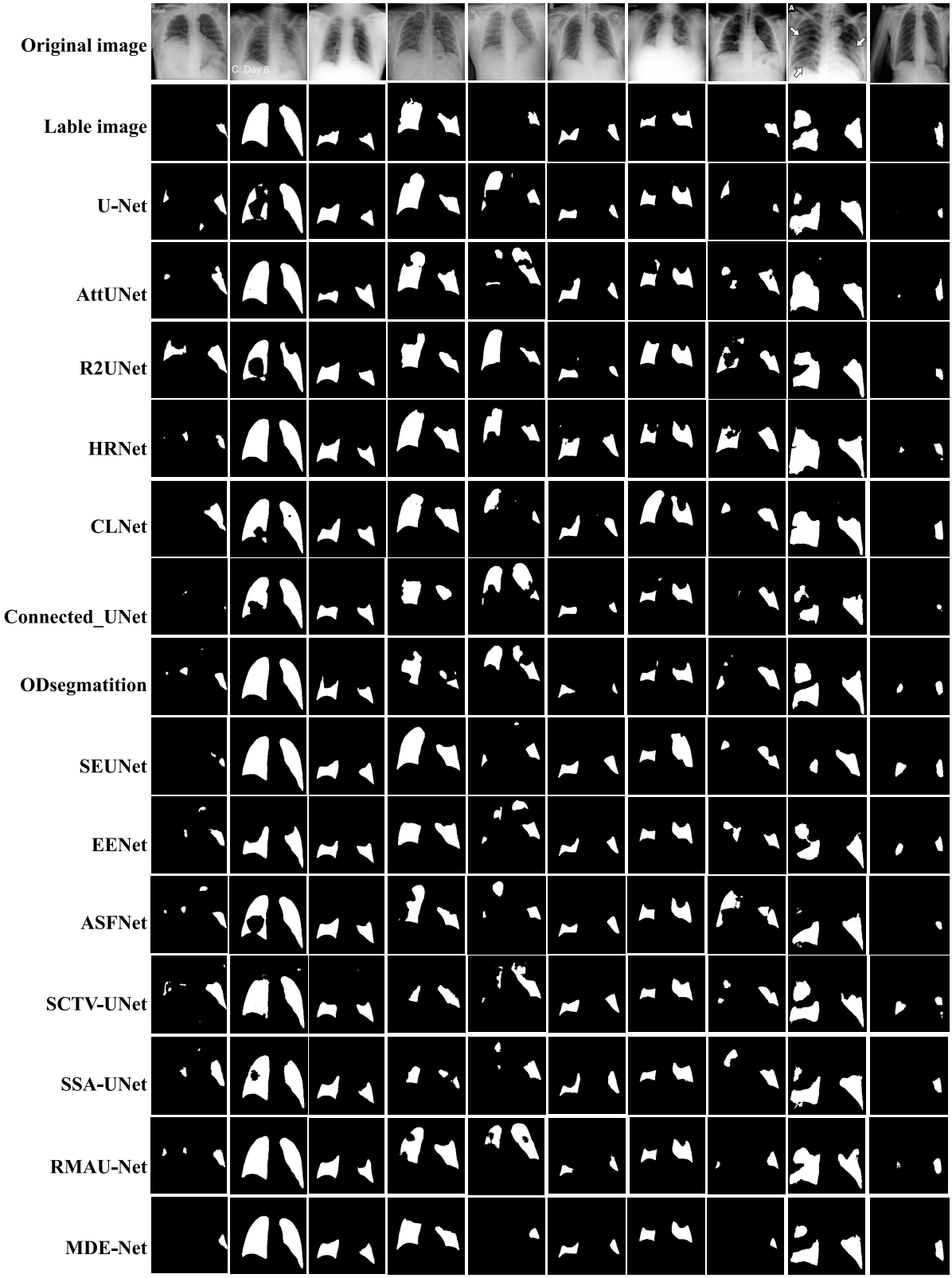

FIGURE 10. Visual segmentation results of different methods on the COVID-19 dataset. First and second rows: original images and their corresponding ground truth. The third to last rows are the results of U-Net, AttUNet, R2UNet, HRNet, CLNet, Connected_UNet, ODsegmatition, SEUNet, EENet, ASFNet, SCTV-UNet, SSA-UNet, RMAU-Net and MDE-Net.

TABLE 2. Comparison experiment on the COVID-19 dataset.

Method	Dice(%)	Mcc(%)	Accuracy(%)	Jaccard(%)
U-Net [5]	81.20	78.34	95.48	68.55
AttUNet [36]	82.50	79.83	95.40	70.42
R2UNet [37]	81.11	78.27	95.20	68.46
HRNet [38]	80.65	77.90	94.71	67.82
CLNet [39]	81.89	79.27	95.20	69.59
Connected_UNet [40]	81.79	79.05	95.48	69.39
ODsegmatiton [41]	76.28	72.77	95.53	61.89
SEUNet [42]	77.18	73.86	95.25	63.13
EENet [43]	80.58	77.62	95.12	67.67
ASFNet [44]	81.94	79.20	95.47	69.65
SCTV-UNet [45]	81.10	78.19	94.76	68.39
SSA-UNet [46]	81.08	78.18	94.69	68.36
RMAU-Net [47]	80.80	77.86	95.29	68.00
MDE-Net	83.43	79.92	95.53	70.93

TABLE 3. Parameters and computational efficiency of different methods on the ISIC-18 dataset.

Method	Parameter(M)	Time(ms/step)
U-Net [5]	2.06	269
AttUNet [36]	8.49	452
R2UNet [37]	16.83	620
HRNet [38]	27.28	610
CLNet [39]	7.72	292
Connected_UNet [40]	3.17	382
ODsegmatiton [41]	19.82	365
SEUNet [42]	1.87	255
EENet [43]	29.98	339
ASFNet [44]	5.63	298
SCTV-UNet [45]	1.13	233
SSA-UNet [46]	2.33	290
RMAU-Net [47]	1.14	246
MDE-Net	5.96	687

complexity comes with a performance sacrifice, as these models tend to achieve lower accuracy in lesion detection. In contrast, models like AttUNet and R2UNet benefit from the integration of residual modules, which enhance feature representation but also result in increased training time and parameter requirements. Similarly, HRNet leverages repeated multi-scale fusion to increase its powerful performance at the expense of computing resources. Despite requiring more training time, our model stands out by providing higher detection accuracy. This finding highlights the need for a balanced approach in model selection, weighing computational requirements with the ultimate goal of achieving accurate and reliable segmentation results.

F. ABLATION STUDIES

To gain a deeper understanding of HCFE block, MPcsSE block and SE-MultiResConv block, we used ablation experiments to analyze the respective contributions of these modules. First, build the original U-Net architecture as our base framework. Then integrate the above modules one by one. Finally, Dice, Mcc, Accuracy and Jaccard were used to evaluate the performance. Table. 4 and Fig. 11 give quantitative and visual results on the ISIC2018 dataset, providing a comprehensive overview of our findings.

1) EFFICACY OF HCFE BLOCK

First, we integrated HCFE block into the base framework, and the third rows of Fig. 11 show a comprehensive visualization. This HCFE block significantly improves the segmentation accuracy of the underlying U-Net framework, which is able to capture a wider range of global feature images. It can be confirmed from Table. 4 that the integration of HCFE block into the Baseline framework (Baseline + HCFE) has shown some improvement in all assessment indicators. It is worth noting that compared with Baseline, Dice, Mcc, Accuracy and Jaccard scores improved from 82.77%, 76.71%, 90.67%, 71.24 to 84.27%, 78.93%, 91.02%, 73.61%, increased by 1.50%, 2.22%, 0.35% and 2.37% respectively. The main reason is the introduction of HCFE block as an encoder, which not only adaptively adjusts convolution kernel parameters according to the characteristics of input data, so as to capture complex patterns and structures in images more accurately, but also allows the model to capture local details while taking into account global context information by expanding the receptive field.

2) EFFICACY OF MPcsSE BLOCK

In the second phase, we incorporated MPcsSE block into the encoder and decoder connection process to improve the model's adaptive ability to filter out more important information. As shown in Table. 4, the segmentation performance is basically equal to that of U-Net. In addition, a more detailed evaluation of the segmentation results is provided through the visual representation in the fourth line of Fig. 11.

3) EFFICACY OF SE-MultiResConv BLOCK

Third, we introduce SE-MultiResConv block into the U-Net basic architecture, expressed as (Baseline + SE-MultiResConv), in order to evaluate its effectiveness. Because the decoder of the basic network uses simple deconvolution and upsampling to capture local information, it cannot accurately mine the small information of the feature image. As shown in the fifth row of Fig. 11 and the comprehensive analysis shown in Table. 4, with the enhancement of SE-MultiResConv block, the network can show greater efficiency and accuracy in processing complex images. The SE attention mechanism is used to dynamically recalibrate the channel importance in the feature map, allowing the model to focus on the features that are most critical to the segmentation task, while ignoring unimportant background noise. The multi-scale convolutional residual module helps the network to better capture the subtle structural changes in the image by fusing the feature information of different scales, and further improves the richness and accuracy of feature extraction.

4) EFFICACY OF FUSION MODULE

Finally, in order to effectively convey context information, we combined HCFE block, MPcsSE block and

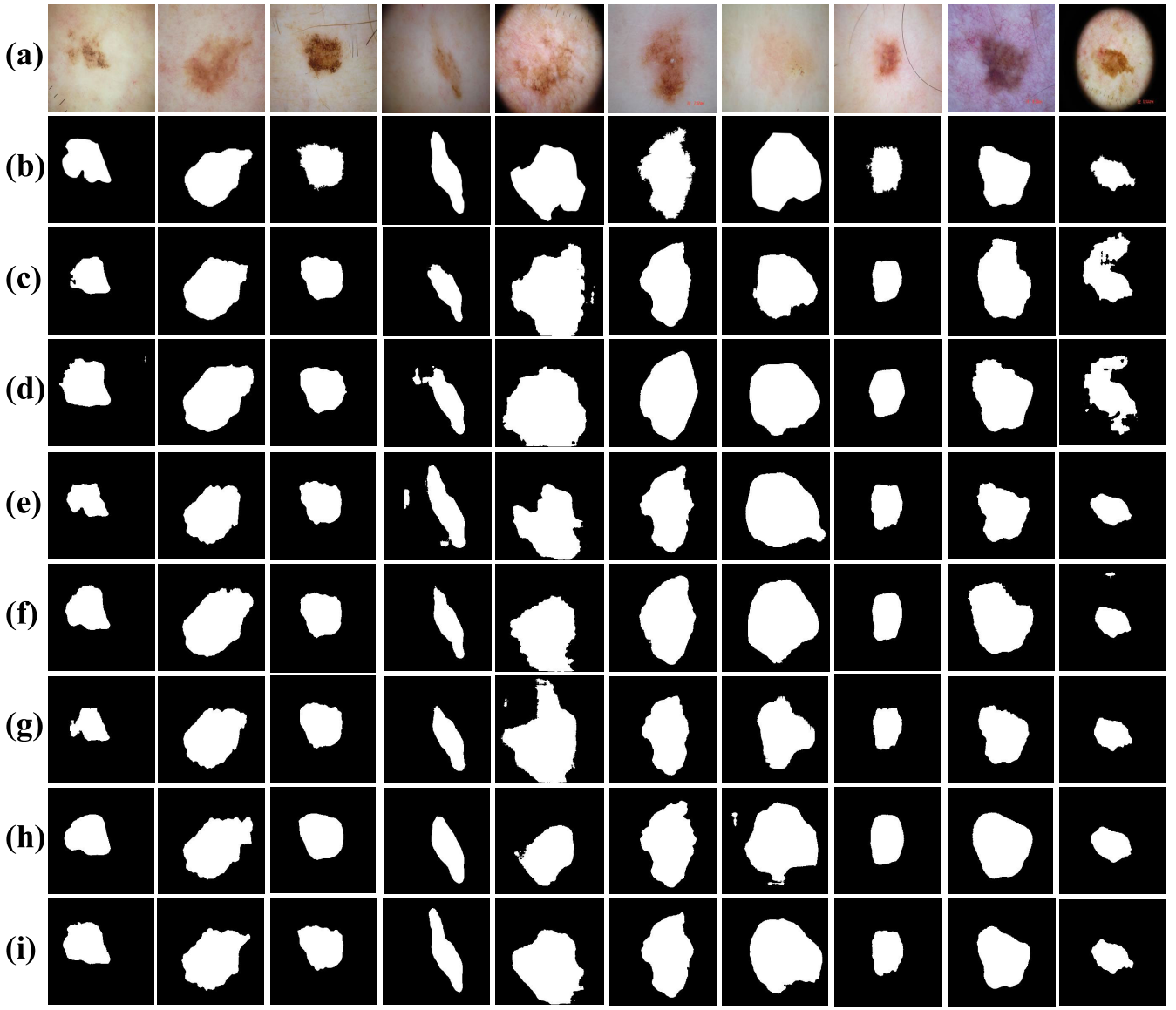

FIGURE 11. Visualization results of ablation studies. (a-b) original images and their corresponding ground truth. (c) Baseline + HCFE. (d) Baseline + MPcsSE. (e) Baseline + SE-MultiResConv. (f) Baseline + HCFE + MPcsSE. (g) Baseline + MPcsSE + SE-MultiResConv. (h) Baseline + HCFE + SE-MultiResConv. (i) Baseline + HCFE + MPcsSE + SE-MultiResConv.

TABLE 4. Ablation experiments on the ISIC-18 dataset.

Method	Dice(%)	Mcc(%)	Accuracy(%)	Jaccard(%)
Baseline	82.77	76.71	90.67	71.24
Baseline+HCFE	84.27	78.93	91.02	73.61
Baseline+MPcsSE	82.75	76.50	90.73	71.23
Baseline+ SE-MultiResConv	85.22	79.80	90.98	75.03
Baseline+ HCFE+MPcsSE	85.03	79.41	90.71	74.52
Baseline+MPcsSE+ SE-MultiResConv	85.65	80,64	91.08	75.70
Baseline+HCFE+ SE-MultiResConv	85.35	80.37	91.01	75.24
Baseline+HCFE+ MPcsSE+SE-MultiResConv	86.63	81.78	91.59	76.98

SE-MultiResConv block to design a fusion module (Baseline + HCFE + MPcsSE + SE-MultiResConv). The combination of the three can better capture the multi-scale information of the image and effectively integrate the feature information of different levels, which can help the model identify the target boundary, small structure and occlusion more accurately. As you can see from the last row of Fig. 11, our approach captures relatively complete global information and finely segmented results compared to the Baseline network. As shown in Table. 4, our approach shows dramatic improvements in Dice, Mcc, Accuracy, and Jaccard scores, with improvements of 3.86%, 5.07%, 0.92%, and 5.47%, respectively, compared to Baseline networks. As can be seen from the images and statistical results, each component in our model exhibits a unique performance, and combining these components together can obtain the optimal segmentation results.

G. LIMITATION

The proposed method offers significant performance benefits, particularly in improving the accuracy of image and data segmentation tasks. However, there remain certain limitations that need be addressed for broader appicability. A major challenge lies in the method's computational efficiency, as it requires substantial processing time, especially when handling large datasets or in scenarios that demand real-time responses. To overcome this obstacle, we will reduce processing time and increase overall computing speed in several ways, including algorithmic optimization, parallel computing techniques, and leveraging hardware acceleration. Moreover, we intend to explore new segmentation algorithms and innovative architectures to further enhance the method's capabilities. By striking a balance between improving computational performance and maintaining high segmentation accuracy, we seek to broaden the method's scope of application, making it more practical and impactful across diverse scenarios.

IV. CONCLUSION

In order to solve the long-standing problems in the field of medical image processing, the limitations of global feature extraction and the problems of fuzzy and incomplete image details. We develop a novel multi-scale feature extraction and attention mechanism combined network. The elaborate design of this network architecture not only shows the fusion application of cutting-edge technologies, but also opens up a new path for the accurate diagnosis of diseases such as skin cancer. As the entrance of MDE-Net, the encoder module is responsible for the initial processing of the input medical image, and gradually abstracts the low-level to high-level features of the image through the multi-layer convolutional neural network structure. In this process, we introduced HCFE. This structure allows the convolution layer to see a wider range of image regions by increasing the ''receptive field'' of the convolution kernel, that is, while keeping the number of parameters unchanged, so as to effectively capture the context information in the image. The intermediate connection extraction module is one of the cores of MDE-Net, which cleverly connects encoder and decoder to realize the deep transmission and fusion of features. In this module, the proposal of MPcsSE is a highlight. By dynamically adjusting the importance between different channels, MPcsSE effectively reduces the interference of noise and redundant information, while enhancing the representation of key features. This adaptive feature selection mechanism enables the model to extract global features more accurately in the complex and changeable medical image environment, which lays a solid foundation for subsequent analysis and diagnosis. The decoder module is responsible for decoding the advanced features extracted by the encoder and the intermediate connection extraction module step by step, and restoring the segmentation map with similar size to the original image. To ensure global consistency while finely restoring image detail, MDE-Net has designed an innovative fusion structure that combines the SE attention mechanism with a multi-scale convolutional residual module. The SE attention mechanism enhances the representation of useful features and suppresses unimportant features by explicitly modeling the interdependencies between channels. The multi-scale convolutional residuals module captures multi-level details in images through convolution kernel of different scales, and relieves the training problem of deep networks through residuals connection. The combination of the two makes MDE-Net able to recover the tiny structure in the image and improve the segmentation accuracy while maintaining the consistency of the global feature. Through a comprehensive evaluation of ISIC18 datasets and COVID-19 datasets, MDE-Net demonstrated significant performance benefits. Compared with other contemporary state-of-the-art segmentation technologies, MDE-Net has achieved significant improvements in several evaluation indicators. It not only improves the accuracy of segmentation, but also enhances the model's ability to capture image details and complex structures. This result validates the rationality and effectiveness of MDE-Net design. It also provides a new idea and direction for the development of medical image processing.

In future research, our team will aim to further advance the field of medical image segmentation and collaborate with some leading hospitals. It mainly includes the automatic recognition and analysis of colorectal images and the detection and evaluation of pituitary tumor lesions. At present, we have systematically built a high-quality, large-scale medical image database, which not only covers a wealth of medical image samples, but also lays a solid foundation for the subsequent algorithm development and validation. Meanwhile, we are deeply aware of the limitations of current models in terms of computational efficiency, which directly affects the immediate application potential of the algorithm in real-world clinical Settings. Therefore, optimizing and improving computing efficiency has become one of the indispensable core directions of our future research.

REFERENCES

[1] L. Qu, Q. Jin, K. Fu, M. Wang, and Z. Song, ''Rethinking deep active learning for medical image segmentation: A diffusion and angle-based framework,'' *Biomed. Signal Process. Control*, vol. 96, Oct. 2024, Art. no. 106493.

[2] T. Dang, T. T. Nguyen, J. McCall, E. Elyan, and C. F. Moreno-García, ''Two-layer ensemble of deep learning models for medical image segmentation,'' *Cognit. Comput.*, vol. 16, no. 3, pp. 1141–1160, May 2024.

[3] H. Messaoudi, A. Belaid, D. B. Salem, and P.-H. Conze, ''Cross-dimensional transfer learning in medical image segmentation with deep learning,'' *Med. Image Anal.*, vol. 88, Aug. 2023, Art. no. 102868.

[4] E. Shelhamer, J. Long, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[5] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Munich, Germany. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[6] M. Xu, Q. Ma, H. Zhang, D. Kong, and T. Zeng, "MEF-UNet: An end-to-end ultrasound image segmentation algorithm based on multi-scale feature extraction and fusion," *Computerized Med. Imag. Graph.*, vol. 114, Jun. 2024, Art. no. 102370.

[7] C. Zhao, W. Lv, X. Zhang, Z. Yu, and S. Wang, "MMS-Net: Multi-level multi-scale feature extraction network for medical image segmentation," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105330.

[8] Y. Feng, X. Zhu, X. Zhang, Y. Li, and H. Lu, "PAMSNet: A medical image segmentation network based on spatial pyramid and attention mechanism," *Biomed. Signal Process. Control*, vol. 94, Aug. 2024, Art. no. 106285.

[9] L. P. M. Ullah, A. Vats, F. A. Cheikh, K. G. Santhosh, and M. S. Nair, "EfficientPolypSeg: Efficient polyp segmentation in colonoscopy images using EfficientNet-B5 with dilated blocks and attention mechanisms," *Biomed. Signal Process. Control*, vol. 93, Jul. 2024, Art. no. 106210.

[10] T. M. Khan, S. S. Naqvi, and E. Meijering, "ESDMR-Net: A lightweight network with expand-squeeze and dual multiscale residual connections for medical image segmentation," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 107995.

[11] M. Mubashar, H. Ali, C. Grönlund, and S. Azmat, "R2U++: A multiscale recurrent residual U-net with dense skip connections for medical image segmentation," *Neural Comput. Appl.*, vol. 34, no. 20, pp. 17723–17739, Oct. 2022.

[12] J. Selvaraj, S. Umapathy, and N. A. Rajesh, "Artificial intelligence based real time colorectal cancer screening study: Polyp segmentation and classification using multi-house database," *Biomed. Signal Process. Control*, vol. 99, Jan. 2025, Art. no. 106928.

[13] J. Selvaraj and S. Umapathy, "CRPU-NET: A deep learning model based semantic segmentation for the detection of colorectal polyp in lower gastrointestinal tract," *Biomed. Phys. Eng. Exp.*, vol. 10, no. 1, Jan. 2024, Art. no. 015018.

[14] P. P. Singh, S. Prasad, A. K. Chaudhary, C. K. Patel, and M. Debnath, "Classification of effusion and cartilage erosion affects in osteoarthritis knee mri images using deep learning model," in *Proc. Int. Conf. Comput. Vis. Image Process.*, 2019, pp. 373–383.

[15] M. J. Khan and P. P. Singh, "Advanced road extraction using CNN-based U-net model and satellite imagery," *e-Prime-Adv. Electr. Eng., Electron. Energy*, vol. 5, Sep. 2023, Art. no. 100244.

[16] Y. Zhang, X. Zhang, and W. Zhu, "ANC: Attention network for COVID-19 explainable diagnosis based on convolutional block attention module," *Comput. Model. Eng. Sci.*, vol. 127, no. 3, pp. 1037–1058, 2021.

[17] Y. Yin, Z. Han, M. Jian, G.-G. Wang, L. Chen, and R. Wang, "AMSUnet: A neural network using atrous multi-scale convolution for medical image segmentation," *Comput. Biol. Med.*, vol. 162, Aug. 2023, Art. no. 107120.

[18] H. Zhang, X. Zhong, G. Li, W. Liu, J. Liu, D. Ji, X. Li, and J. Wu, "BCU-Net: Bridging ConvNeXt and U-net for medical image segmentation," *Comput. Biol. Med.*, vol. 159, Jun. 2023, Art. no. 106960.

[19] Y. Li, Z. Wang, L. Yin, Z. Zhu, G. Qi, and Y. Liu, "X-Net: A dual encoding–decoding method in medical image segmentation," *Vis. Comput.*, vol. 39, no. 6, pp. 2223–2233, Jun. 2023.

[20] J. Zheng, H. Liu, Y. Feng, J. Xu, and L. Zhao, "CASF-Net: Cross-attention and cross-scale fusion network for medical image segmentation," *Comput. Methods Programs Biomed.*, vol. 229, Feb. 2023, Art. no. 107307.

[21] X. Xie, W. Zhang, X. Pan, L. Xie, F. Shao, W. Zhao, and J. An, "CANet: Context aware network with dual-stream pyramid for medical image segmentation," *Biomed. Signal Process. Control*, vol. 81, Mar. 2023, Art. no. 104437.

[22] B. Zhang, S. Qi, Y. Wu, X. Pan, Y. Yao, W. Qian, and Y. Guan, "Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for segmenting lung tumor from CT images," *Comput. Methods Programs Biomed.*, vol. 222, Jul. 2022, Art. no. 106946.

[23] S. Tyagi and S. N. Talbar, "CSE-GAN: A 3D conditional generative adversarial network with concurrent squeeze-and-excitation blocks for lung nodule segmentation," *Comput. Biol. Med.*, vol. 147, Aug. 2022, Art. no. 105781.

[24] A. Mehrtash, W. M. Wells, C. M. Tempany, P. Abolmaesumi, and T. Kapur, "Confidence calibration and predictive uncertainty estimation for deep medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3868–3878, Dec. 2020.

[25] Y.-Z. Li, Y. Wang, Y.-H. Huang, P. Xiang, W.-X. Liu, Q.-Q. Lai, Y.-Y. Gao, M.-S. Xu, and Y.-F. Guo, "RSU-Net: U-net based on residual and self-attention mechanism in the segmentation of cardiac magnetic resonance images," *Comput. Methods Programs Biomed.*, vol. 231, Apr. 2023, Art. no. 107437.

[26] J. Selvaraj and A. K. Jayanthy, "Design and development of artificial intelligence-based application programming interface for early detection and diagnosis of colorectal cancer from wireless capsule endoscopy images," *Int. J. Imag. Syst. Technol.*, vol. 34, no. 2, Mar. 2024, Art. no. e23034.

[27] J. Selvaraj and A. K. Jayanthy, "Automatic polyp semantic segmentation using wireless capsule endoscopy images with various convolutional neural network and optimization techniques: A comparison and performance evaluation," *Biomed. Eng., Appl., Basis Commun.*, vol. 35, no. 6, Dec. 2023, Art. no. 2350026.

[28] Y. Yang, C. Feng, and R. Wang, "Automatic segmentation model combining U-net and level set method for medical images," *Expert Syst. Appl.*, vol. 153, Sep. 2020, Art. no. 113419.

[29] A. Selvaraj and E. Nithiyaraj, "CEDRNN: A convolutional encoder–decoder residual neural network for liver tumour segmentation," *Neural Process. Lett.*, vol. 55, no. 2, pp. 1605–1624, Apr. 2023.

[30] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, Mar. 2024.

[31] Q. Zhu, "On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset," *Pattern Recognit. Lett.*, vol. 136, pp. 71–80, Aug. 2020.

[32] J. O. B. Diniz, J. L. Ferreira, P. H. B. Diniz, A. C. Silva, and A. C. Paiva, "A deep learning method with residual blocks for automatic spinal cord segmentation in planning CT," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103074.

[33] X. Fan, J. Zhou, X. Jiang, M. Xin, and L. Hou, "CSAP-UNet: Convolution and self-attention paralleling network for medical image segmentation with edge enhancement," *Comput. Biol. Med.*, vol. 172, Apr. 2024, Art. no. 108265.

[34] H. Abdel-Nabi, M. Z. Ali, and A. Awajan, "A multi-scale 3-stacked-layer coned U-net framework for tumor segmentation in whole slide images," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105273.

[35] K. Hu, Y. Zhu, T. Zhou, Y. Zhang, C. Cao, F. Xiao, and X. Gao, "DSC-Net: A novel interactive two-stream network by combining transformer and CNN for ultrasound image segmentation," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 5030012.

[36] S. Lian, Z. Luo, Z. Zhong, X. Lin, S. Su, and S. Li, "Attention guided U-net for accurate iris segmentation," *J. Vis. Commun. Image Represent.*, vol. 56, pp. 296–304, Oct. 2018.

[37] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-net (R2U-net) for medical image segmentation," 2018, *arXiv:1802.06955*.

[38] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5693–5703.

[39] Z. Zheng, Y. Wan, Y. Zhang, S. Xiang, D. Peng, and B. Zhang, "CLNet: Cross-layer convolutional neural network for change detection in optical remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 247–267, May 2021.

[40] A. Baccouche, B. Garcia-Zapirain, C. C. Olea, and A. S. Elmaghraby, "Connected-UNets: A deep learning architecture for breast mass segmentation," *NPJ Breast Cancer*, vol. 7, no. 1, p. 151, Dec. 2021.

[41] L. Wang, J. Gu, Y. Chen, Y. Liang, W. Zhang, J. Pu, and H. Chen, "Automated segmentation of the optic disc from fundus images using an asymmetric deep learning network," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107810.

[42] L.-Y. Jiang, C.-J. Kuo, O. Tang-Hsuan, M.-H. Hung, and C.-C. Chen, "SE-U-Net: Contextual segmentation by loosely coupled deep networks for medical imaging industry," in *Proc. Asian Conf. Intell. Inf. Database Syst.* Cham, Switzerland: Springer, 2021, pp. 678–691.

[43] L. Wang, M. Shen, C. Shi, Y. Zhou, Y. Chen, J. Pu, and H. Chen, "EE-Net: An edge-enhanced deep learning network for jointly identifying corneal micro-layers from optical coherence tomography," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103213.

[44] J. Chen, Y. Jiang, L. Luo, and W. Gong, "ASF-Net: Adaptive screening feature network for building footprint extraction from remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4706413.

[45] X. Liu, Y. Liu, W. Fu, and S. Liu, "SCTV-UNet: A COVID-19 CT segmentation network based on attention mechanism," *Soft Comput.*, pp. 1–11, Mar. 2023, doi: 10.1007/s00500-023-07991-7.
[46] S. Jiang, X. Chen, and C. Yi, "SSA-UNet: Whole brain segmentation by U-net with squeeze-and-excitation block and self-attention block from the 2.5D slice image," *IET Image Process.*, vol. 18, no. 6, pp. 1598–1612, May 2024.
[47] L. Jiang, J. Ou, R. Liu, Y. Zou, T. Xie, H. Xiao, and T. Bai, "RMAU-Net: Residual multi-scale attention U-net for liver and tumor segmentation in CT images," *Comput. Biol. Med.*, vol. 158, May 2023, Art. no. 106838.

XIAOKANG DING received the M.S. and Ph.D. degrees in forest engineering from Beijing Forestry University. She is currently an Associate Professor with Quzhou University. Her research interests include machine vision and image recognition.

LING DONG was born in Zhoushan, China, in 2004. She is currently pursuing the degree with the School of Mechanical Engineering, Quzhou University. Her current research interests include deep learning and machine learning.

YINGYU JI received the master's degree in control engineering from Zhejiang University of Technology, Hangzhou, China. She is currently a Senior Experimentalist with Quzhou University. Her research interests include deep learning and numerical simulation of fluid machinery.

KE'ER QIAN was born in Zhejiang, China, in 2002. She received the bachelor's degree majoring in mechanical design and manufacturing and automation from Quzhou University, China, in 2022. Her current research interest includes deep learning image segmentation.