RESEARCH ARTICLE

MSR-Net: Multi-Scale Residual Network Based on Attention Mechanism for Pituitary Adenoma MRI Image Segmentation

QILE ZHANG[1], XIAOLIANG JIANG[2], XIUQING HUANG[1], AND CHUN ZHOU[1]

[1]Department of Rehabilitation, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou People's Hospital, Quzhou 324000, China
[2]College of Mechanical Engineering, Quzhou University, Quzhou 324000, China

Corresponding authors: Xiaoliang Jiang (jxl_swjtu@163.com) and Xiuqing Huang (13059779571@163.com)

ABSTRACT Accurate segmentation of pituitary adenoma lesions is essential for effective diagnosis and treatment planning. However, traditional algorithms struggle with this task due to the variability in lesion shape, position, and the presence of extensive background areas. In response to the above challenges, we develop a segmentation framework for pituitary adenoma that integrates multi-scale residual, channel attention and spatial attention mechanisms. In the encoding stage of MSR-Net, a multi-scale residual block is introduced to enhance the ability of channel information extraction across varying scales. In the decoding phase, we construct two paths: the introduction of channel attention allows for obtaining the nuanced weighting of the response degree of each channel to key information, and the spatial attention is utilized to extract the global dependence of features to ease the interference of complex background on segmentation performance. When tested on the constructed original pituitary adenoma database, the specificity, IoU, Mcc, and Dice of MSR-Net reached 99.74%, 80.87%, 89.12%, and 89.34%, which were 0.16%, 5.96%, 3.76%, and 3.76% higher than traditional U-Net. Furthermore, we embarked extensive ablative studies on the original dataset to dissect and evaluate the efficacy of each key modules integrated into our architectural framework. Finally, we expanded upon the original dataset to create an enhanced database, and the objective evaluation index further verifies the superiority and advanced nature of MSR-Net. Compared with other most advanced segmentation approaches, MSR-Net shows superior segmentation effect and robustness, as well as great potential in clinical application, which provides an important reference for the future research and development of medical image segmentation.

INDEX TERMS Image segmentation, U-Net, pituitary adenoma, multi-scale residual, attention mechanism.

I. INTRODUCTION

Pituitary adenoma stands as one of the prevalent intracranial tumors encountered in clinical practice, and most of its MR images are characterized by high noise, artifacts, and similarity to surrounding tissues. However, due to the complexity of the anatomy around the pituitary gland, doctors are often subject to subjective experience during the diagnosis process, which may lead to misdiagnosis and subsequent treatment

The associate editor coordinating the review of this manuscript and approving it for publication was Yeon-Ho Chung.

challenges. Currently, recent advances in clinical medical imaging data and artificial intelligence offer promising solutions to address these diagnostic barriers. By using various image processing algorithms, the interference of subjective factors can be minimized. Firstly, it greatly improves the accuracy of segmentation and can more accurately locate the lesion area. Secondly, analysis based on image processing algorithms can provide doctors with more quantitative information, so this technology helps to formulate more precise diagnosis and treatment plan. In summary, integrating image processing algorithms into medical practice has great
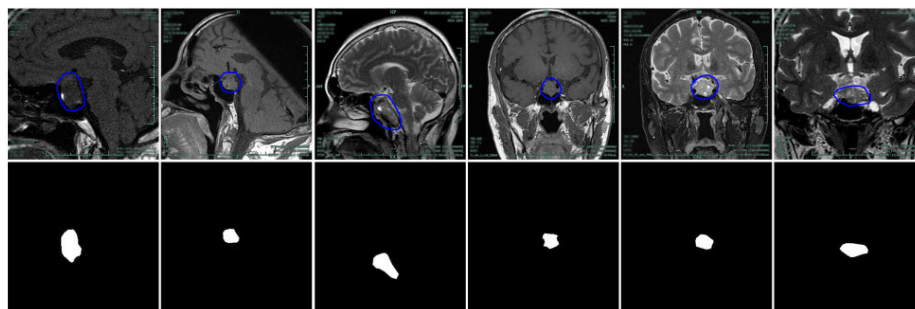
FIGURE 1. MR images of pituitary adenoma and corresponding gold standard.

potential to improve diagnostic accuracy and patient outcomes for pituitary adenoma.

Although machine vision-based medical image segmentation technology has achieved good results in theoretical research and practical application, it still faces the following difficulties: 1) interfered by external factors such as imaging system, lighting conditions, instruments and equipment; 2) the medical lesion tissue also showed characteristics such as low resolution, high noise, variable location, and no structural information, as shown in Figure 1. However, in the face of the above challenges, deep-learning technology has become a powerful tool for medical image-assisted diagnosis [1], [2]. Compared with traditional segmentation frameworks [3], [4], deep-learning offers a distinct advantage by autonomously uncovering and assimilating higher-level image features directly from training data. This inherent capability notably reduces the need for manual feature refinement and enables seamless end-to-end image processing within neural network structures. With the emergence of fully neural network (FCN) [5], automated medical image analysis can achieve near-radiologist level performance. Specifically, by transforming fully connected layers into convolutional layers, FCN can capture low-level and high-level features within images. In addition, the up-sampling method is used for pixel-level prediction, which can assign a category label or segmentation mask to each pixel. Subsequently, in view of the characteristics of biomedical images, Ronneberger et al. [6] introduced U-Net architecture to restore spatial resolution in input images. Among them, the encoder gradually down-samples feature spatial dimensions, while the decoder repairs the detail and spatial resolution, and the information is transmitted between the encoder and the decoder through a ship connection.

In recent years, more and more researchers have begun to pay attention to image semantic segmentation of pixel-level labels, and obtain segmentation performance comparable to existing methods. Among them, Tang et al. [7] introduced a groundbreaking end to end network specifically designed to address the uncertainty inherent in segmentation tasks. Specifically, initially, a rough segmentation module was utilized to generate initial segmentations alongside uncertainty graphs. Then, an uncertainty guidance block was designed to obtain an uncertainty map in the end-to-end network. Thirdly,

the module contains a double attention block, which was further refined by a feature refinement module to improve the final segmentation quality. Concurrently, Sun et al. [8] constructed a double path network based on DeepLabV3+ to tackle the detection and segmentation of thyroid nodules. This mechanism orchestrated a synergistic interplay between different paths, enabling the model to effectively leverage both regional and shape-based cues for segmentation. Similarly, Zhang et al. [9] introduced a pioneering retinal vessel segmentation methodology, drawing upon the innovative M-Net architecture as its foundation. Firstly, this method encompassed a double attention mechanism, operating at both channel and spatial levels, to mitigate the impact of noise in the input data. Secondly, self-focused mechanisms from Transformer were incorporated into skip connections, effectively re-coding features and enhancing feature representations. Finally, to harness the unique characteristics of each layer, a weighted side layer was presented to optimize the utilization of information from different network depths. Meanwhile, Fu et al. [10] presented a pioneering automated segmentation technique meticulously crafted for the unique challenges posed by cardiac MRI images. Firstly, their approach combined the strengths of CNNs for detailed feature acquisition and spatial encoding, giving full play to the ability of convolutional in detail capture. Additionally, they harness the power of Transformer models to capture remote dependencies among features across various scales, further enriching the segmentation process with contextual information.

Subsequently, many scholars have conducted a lot of research on the exploration and utilization of multi-scale features [11], [12]. For example, considering the diverse shapes and uneven locational distribution of hybridoma cells, Lu et al. [13] proposed a RA-UNet segmentation network to address these challenges. This framework performs multi-scale splicing of encoding features using a pyramid structure, which allows for the capture of information at various levels of detail. In addition, it employs a non-local focus mechanism, which improves the overall performance and accuracy of cell segmentation. Yuan et al. [14] introduced the McNfm-UNet, a sophisticated segmentation framework designed to enhance the utilization of feature information and address the gradient vanishing problem. This framework

employed multi-layer fusion and multi-scale feature mapping at each stage of the network, and fused with higher-level abstractions from later layers to maximize the use of all available feature information. Saeed et al. [15] introduced an innovative multi-scale feature fusion module designed to enhance the robustness and discriminative power of feature representations in image segmentation tasks. Xu et al. [16] developed an end-to-end model called MEF-UNet, which includes a sophisticated multi-scale feature fusion module integrated into the decoder segment. This approach not only addresses the challenges associated with scaling variations in feature maps, but also enhances the ability to capture and combine complex details and broader contextual information in images.

Inspired by the above algorithms, this article proposes a framework for pituitary adenoma segmentation by combining multi-scale residual network and attention mechanism which can guarantee high segmentation accuracy without increasing computation time. This makes the network not only make full use of image features, but also adaptively adjust and fuse channel information, thus further strengthen the feature expression ability. The key innovations of our approach are given as:

1) We design an end-to-end pituitary tumor segmentation network (MSR-Net), which achieves an optimal balance between segmentation accuracy and computational efficiency.

2) Our methodology integrates a multi-scale residual module to obtain shallow multi-dimensional feature information of channel matrix, which enhances the ability of extracting feature information of channel across varying scales, and overcomes the problem of extracting single channel scale in deep-learning.

3) Two decoding paths are designed, in which the channel attention is used to obtain the relative importance of each channel's response to key information, and the spatial attention is utilized to extract the global dependence of features. Meanwhile, the feature maps of the two paths cooperate with each other to obtain accurate feature re-presentation.

II. METHODS

A. OVERVIEW OF MSR-NET

At present, the segmentation of pituitary adenoma is faced with the following challenges: low resolution, high noise, variable location, and no structural information. The standard U-Net architecture employs two 3×3 convolutional layers, ReLU activations, max-pooling, up-sampling, and skip connections, which may not capture the intricate details and varying scales of lesion features, especially in cases with heterogeneous and subtle variations. In response to the above
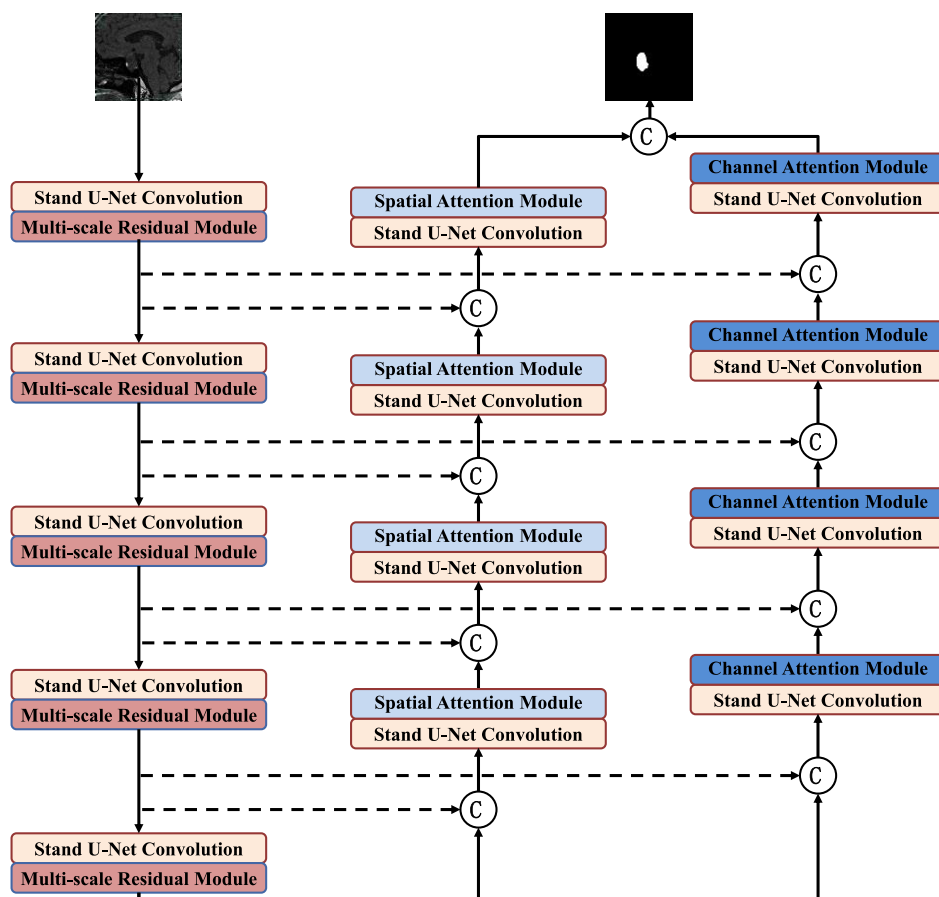

FIGURE 2. Architecture of MSR-Net.

problems, we develop a segmentation framework for pituitary adenoma that integrates multi-scale residual, spatial attention mechanisms and channel attention mechanisms, as illustrated in Figure 2. Specifically, an image is first input into the encoder, which consists of a stand U-Net convolution (3 × 3 convolutional) layer coupled with a multi-scale residual module, aiming to increase the extraction ability of channel information across varying scales. Subsequently, the feature map undergoes refinement through a maximum pooling layer, strategically employed to augment the receptive field. After that, the refined features traverse through two parallel attention decoders, and each step in the decoding path includes a 2×2 up-sampling operation of the feature map inherited from the preceding layer, which is intricately intertwined with the feature map obtained from the encoding path. Finally, the two parallel channel features are connected to realize the adaptive coordination of the two attention modules and complete the pixel-by-pixel prediction of pituitary adenoma lesions.
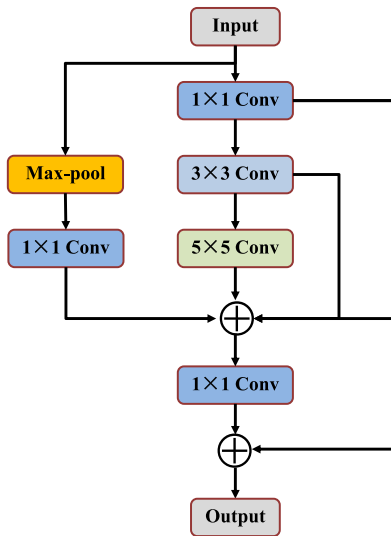

FIGURE 3. Structure of multi-scale residual module.

B. MULTI-SCALE RESIDUAL MODULE
Extensive research has shown the profound impact of network depth on its expressive capacity and learning prowess. However, this augmentation comes at a cost. With the increasing of network depth, the complexity of training network and the elevated loss value also increase than that of the shallow network. Surprisingly, the shallow networks exhibit a stronger expression ability than the deep network. Moreover, the choice of convolution kernel size emerges as a pivotal factor in shaping network performance. Large convolutional kernels cover many features, so it is easy to extract the feature map comprehensively. However, small convolution kernels cover fewer features, so they can extract local details better. Therefore, inspired by the idea of residuals [17], [18], we construct a multi-scale residual block to harness the strengths of different convolutional operations, as shown in Figure 3. The

structure has four distinct branches: 1×1, 3×3, and 5×5 convolutions, alongside max-pooling. Each branch specializes in capturing features across varying spatial scales, enriching the network's capacity to discern patterns at different levels of granularity. Following convolution, strategic steps such as channel reduction through 1×1 convolutions and feature graph normalization are employed to refine and optimize the extracted features. Then, the feature graph is normalized and ReLU is used to improve the nonlinear fitting performance. Finally, to prevent too serious information loss, residual connection is introduced to obtain better network training effect. This architectural element facilitates the seamless integration of earlier layer information with subsequent computations, fostering a more stable and effective training process.

C. CHANNEL ATTENTION MODULE
The attention mechanism, originally pioneered in natural language processing, has since found application in various domains, and then realized pixel-level prediction of images in semantic segmentation tasks [19], [20]. This mechanism plays a pivotal role in filtering information and can discern which areas require heightened attention and consequently assign varying degrees of importance. Among them, channel attention can suppress useless channel features and enhance useful channel features by assigning distinct weights to each channel in the feature map, so that network attention can focus on critical regions of interest. The channel attention module proposed in our framework is shown in Figure 4.

Firstly, the two-dimensional image of each channel in the feature graph is compressed into a value by global maximum pooling and average pooling, that is, the three-dimensional feature graph is transformed into a one-dimensional vector. Secondly, it passes through a multi-layer perceptron containing two fully connected networks and a ReLU activation. Thirdly, the outputs of the two paths are superimposed by elements, and the result is nonlinear mapped by Sigmoid function to obtain the final channel weight graph. Finally, it is multiplied with the elements of the corresponding channel in feature map I to obtain the recalibrated output feature map I^C. The specific calculation process is as follows:

$$I^C = I \cdot \sigma(W_1\delta(W_0 I_{max}) + W_1\delta(W_0 I_{avg})) \qquad (1)$$

where W_0 and W_1 represent the weights associated with the first and second fully connected layers within multi-layer perceptron, σ and δ denote the Sigmoid and ReLU functions, I_{max} and I_{avg} are the feature map after global maximum pooling and average pooling.

D. SPATIAL ATTENTION MODULE
The channel attention operates by dynamically assigning weights to individual channel of the input feature graph to determine "which graph" has more feature expression ability, while the spatial attention, as a complementary mechanism to channel attention, pays more attention to "where" of the feature graph is important information, and obtains more contextual information from local features. Therefore,
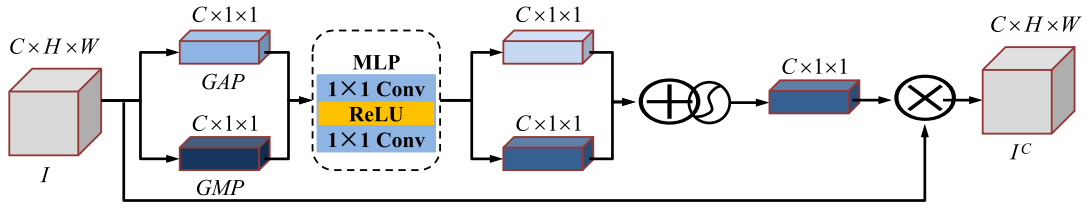
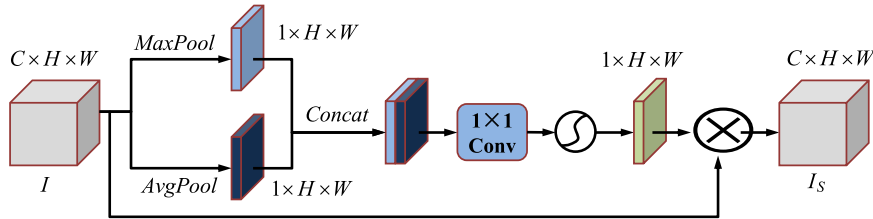**FIGURE 4.** Structure of channel attention module.

FIGURE 5. Structure of spatial attention module.

to extract the global dependency of features and alleviate the interference of complex background on segmentation performance, the spatial attention module is strategically integrated into one of the decoder components, as shown in Figure 5.

Firstly, to aggregate the feature information of each pixel, the average and maximum pooling of input features I are performed. Each of these pooling methods serves a distinct purpose in capturing different facets of the feature map. Among them, average pooling focuses on the extraction of overall features in the feature map, and maximum pooling focuses on the extraction of detailed features within the feature map. Following this, the resulting feature graphs from both pooling operations are combined along the channel dimension and the spatial weights of the feature graphs are mapped to (0-1) through the 1×1 convolution layer and Sigmoid function. Finally, these weighted coefficients are applied to the original input features through a multiplication process to get the output features I^S. The formula for this process is described as:

$$I^S = I \cdot \sigma(f^{1 \times 1}[I_{max}, I_{avg}]) \quad (2)$$

where $f^{1 \times 1}$ is the convolution operation with kernel size 1.

E. LOSS FUNCTION
In the realm of deep-learning, training networks involves continuously evaluating the variance between the predicted outcome and the underlying truth value by using a loss function. This loss function serves as a guidepost, quantifying the disparity between predicted and actual values, which is then utilized to iteratively adjust the model's parameters via back-propagation until convergence is achieved. However, due to the presence of more background areas in MRI images and the small area occupied by pituitary tumors, there is a serious class imbalance in the data. Therefore, we incorporate the dice loss function [21], [22] into the training regimen,

which can offer a strategic solution to the class imbalance conundrum and enhance the stability of the model training process. The dice loss function is given as:

$$L_{dice}(y, p) = 1 - \frac{2 \sum_{i=1}^{N} p_i y_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i} \quad (3)$$

where N denotes the number of all pixels, p_i and y_i denote the ground truth and predicted value of pixel I, with values ranging from 0 to 1.

III. EXPERIMENTAL RESULTS
This study was conducted in cooperation with Quzhou People's Hospital, and experts used Labelme annotation tool to manually label the pituitary adenoma database, with a total of 2105 images. Given the inherent variability in image sizes, a standardized resolution of 256×256 was adopted to ensure uniformity across the dataset. In this experiment, the original pituitary adenoma dataset was divided into training, validation, and testing subsets through a random selection process. Specifically, 1400 images were allocated for training, 305 for validation, and 400 for testing. To address the challenge of small sample sizes and imbalanced data distribution, advanced enhancement techniques including random rotation, scaling, panning, cropping, and innovative mosaic enhancement methods were utilized to extend the original dataset. The new enhanced database includes 6312 images, of which 4199 images were allocated for training, 914 images were set aside for validation, 1199 images were earmarked for testing. Table 1 provides a detailed overview of the specifications associated with each dataset. Moreover, all experiments used the windows 10 64-bit operating system, the processor was i7-7700HQ, the memory was 24G, the graphics card was NVIDIA RTX6000. The software

environment adopts Tensorflow 2.0 framework for training on the CUDA10.1 computing platform, and Python3.6 version is programmed. The Adam optimizer is used in the training of MSR-Net, the initial learning rate was set at 0.001, training epoch is set to 200 with a batch size of 16. To uphold the integrity and efficacy of the training process, a termination criterion was established, stipulating that training would cease if the loss function failed to exhibit improvement over a span of 20 consecutive iterations.

TABLE 1. Descriptions of the pituitary adenoma datasets.

Dataset	Number	Training	Validation	Testing
Original	2105	1400	305	400
Enhanced	6312	4199	914	1199

A. EVALUATION METRICS

To verify the validity of multi-scale residual, channel attention and spatial attention fairly and rationally, and to compare with other classical algorithms, four indicators including specificity [23], [24], IoU [25], [26], Mcc [27], [28] and Dice [29], [30] were adopted, which can be given as:

$$specificity = \frac{TN}{TN + FP} \tag{4}$$

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

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

$$Dice = \frac{2TP}{2TP + FN + FP} \tag{7}$$

B. PARAMETER SELECTION

Table 2 presents a comprehensive analysis of the performance metrics obtained by employing various optimizers during the training phase of MSR-Net on the original pituitary adenoma dataset. Among the various optimization algorithms evaluated, although SGD is widely used for machine learning optimization, it has not shown the expected efficiency level in this specific environment. Adamax, Nadam, and RMSprop are optimization algorithms derived from the Adam optimizer, and each with unique modifications to address specific optimization obstacles encountered during model training. It is obvious that they usually show slightly lower performance than Adam when compared to the original Adam optimizers. However, despite their commendable performance, it is discernible that these Adam-derived algorithms typically exhibit a marginally lower performance profile compared to Adam optimizer across various evaluation metrics. Notably, the MSR-Net model attains remarkable specificity, IoU, Mcc and Dice, reaching an impressive 99.74%, 80.87%, 89.12%, and 89.34%, respectively. Therefore, Adam emerges as the optimal choice, which provides a balance of accuracy and robustness for the MSR-Net model.

TABLE 2. Experiments of MSR-Net with different optimizer on the original dataset.

Optimizer	Specificity	IoU	Mcc	Dice
Adamax	0.9973	0.8076	0.8907	0.8929
Nadam	0.9973	0.8016	0.8868	0.8891
RMSprop	0.9969	0.8060	0.8899	0.8918
SGD	0.9942	0.6992	0.8194	0.8212
Adam	**0.9974**	**0.8087**	**0.8912**	**0.8934**

The selection of an appropriate learning rate is a critical aspect of training neural networks, as it directly influences the convergence speed and stability of the optimization process. Table 3 provides a comprehensive analysis of the MSR-Net with different learning rates on the original pituitary adenoma dataset. The experiment encompasses a wide range of learning rates, spanning from 0.0001 to 0.1, to thoroughly explore their impact on segmentation performance. Notably, the specificity, IoU, Mcc, and Dice coefficient metrics are employed to evaluate the effectiveness of each learning rate setting. At a learning rate of 0.0001, the model achieves a specificity of 99.50%, IoU of 73.53%, Mcc of 84.44%, and Dice coefficient of 84.61%.When the learning rate is increased to 0.0005, the segmentation performance of all indicators was significantly improved, with a specificity of 99.73%, IoU of 79.58%, Mcc of 88.29%, and Dice coefficient of 88.53%.With the learning rate increased to 0.001, the optimal performance can be obtained, where the specificity was 99.74%, IoU is 80.87%, Mcc is 89.12%, and Dice coefficient is 89.34%. As the learning rate continued to increase, the segmentation performance began to deteriorate. When the learning rate is 0.1, the performance of the model decreases significantly. The experiment proves that the learning rate of 0.001 is the most suitable choice for training the original pituitary adenoma dataset.

TABLE 3. Experiments of MSR-Net with different learning rate on the original dataset.

Learning rate	Specificity	IoU	Mcc	Dice
0.0001	0.9950	0.7353	0.8444	0.8461
0.0005	0.9973	0.7958	0.8829	0.8853
0.001	**0.9974**	**0.8087**	**0.8912**	**0.8934**
0.005	0.9971	0.7891	0.8788	0.8812
0.01	0.9969	0.7532	0.8546	0.8572
0.05	0.9911	0.5571	0.7083	0.7109
0.1	0.9857	0.3273	0.4737	0.4810

C. RESULTS OF MSR-NET

The MSR-Net underwent meticulous training on the original pituitary adenoma dataset, and its performance was verified on the test set. The number of iterations for training was 200, and the curve of loss and accuracy during the network training phase is provided in Figure 6. Notably, the loss value curve of the training set can decline quickly and smoothly, converge on the basis around 20 iterations, and finally stabilize at about 0.067. In the training process, the loss of the verification
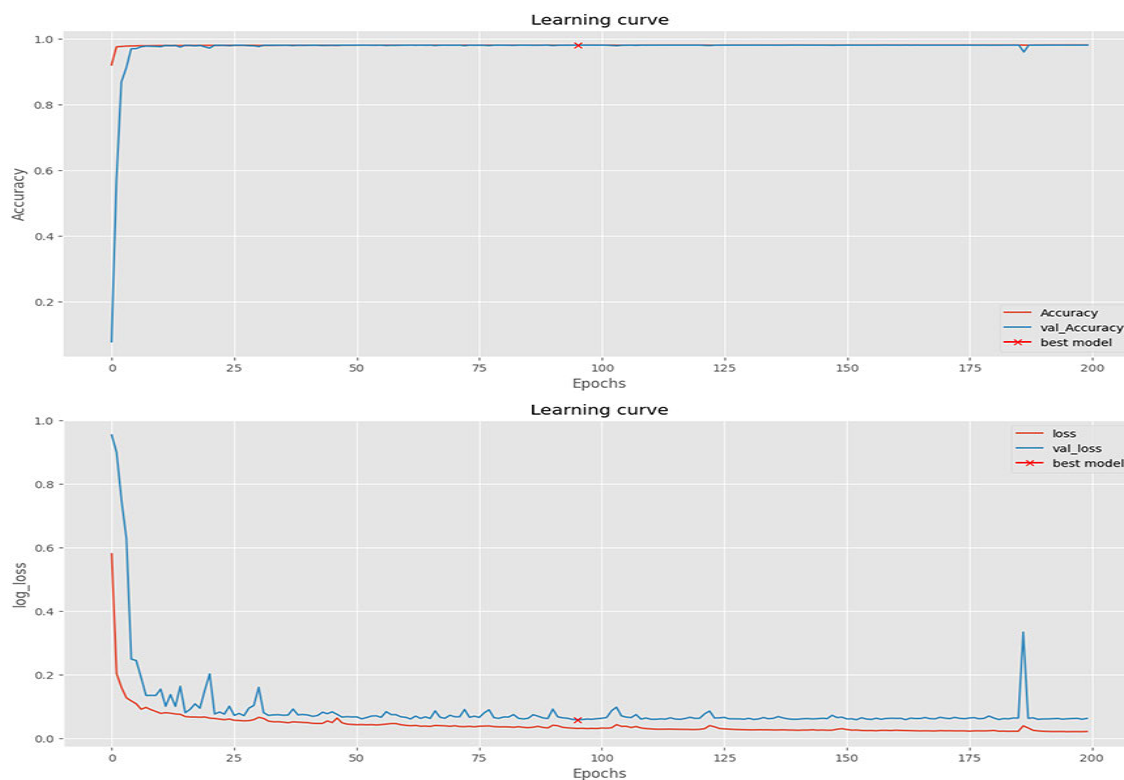
FIGURE 6. Curve evolution of accuracy and loss during training process on the original pituitary adenoma dataset.

set can quickly change with the change of the loss value of the training set. In addition, the accuracy of the validation set could closely follow the change of the accuracy of the training set, which indicated that MSR-Net had strong learning and prediction ability on the pituitary adenoma images, showing good segmentation performance. After the network training, MSR-Net conducts experiment on the test dataset, with the results showcased in Figure 7. Visually, MSR-Net realized accurate segmentation of pituitary adenoma, and the

lesion was obtained with complete shape and smooth edges, effectively excluding different brightness, contrast and other interference in the image, and was not easily affected by weak and pseudo-boundaries.

D. ABLATION STUDIES
To thoroughly assess the efficacy of channel attention (CA) module, spatial attention (SA) module, multi-scale residual (MSR) module in MSR-Net algorithm, U-Net
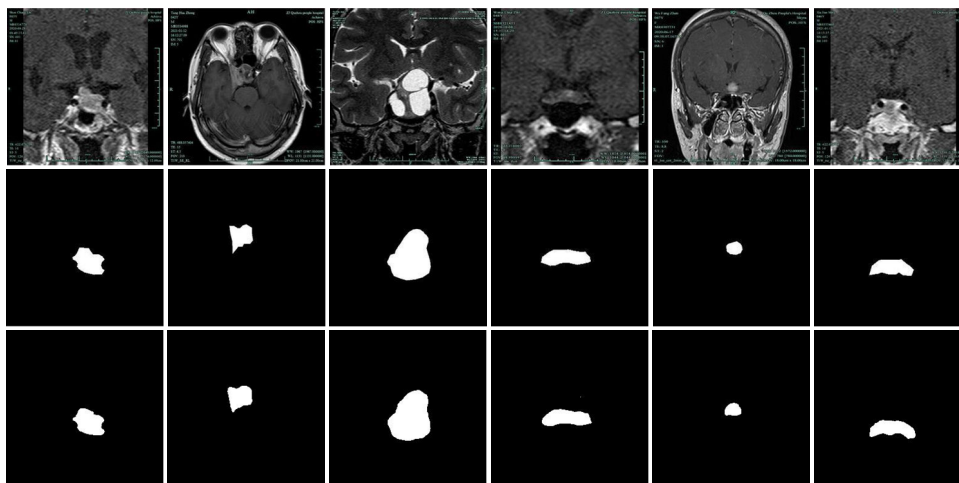

FIGURE 7. Visualization results of MSR-Net on the original pituitary adenoma dataset. The first to last rows are the original images, the corresponding labels, and the segmentation results of MSR-Net.

TABLE 4. Ablation studies of different structures on the original pituitary adenoma dataset.

Method	Specificity	IoU	Mcc	Dice
Baseline (U-Net)	0.9959	0.7491	0.8537	0.8558
Baseline+CA	0.9961	0.7641	0.8634	0.8656
Baseline+SA	0.9963	0.7623	0.8621	0.8644
Baseline+MSR	0.9969	0.7635	0.8626	0.8652
Baseline+CA+SA	0.9966	0.7679	0.8654	0.8679
Baseline+CA+MSR	0.9966	0.7731	0.8690	0.8712
Baseline+ SA +MSR	0.9972	0.7926	0.8814	0.8838
Baseline+CA+SA+MSR (MSR-Net)	**0.9974**	**0.8087**	**0.8912**	**0.8934**
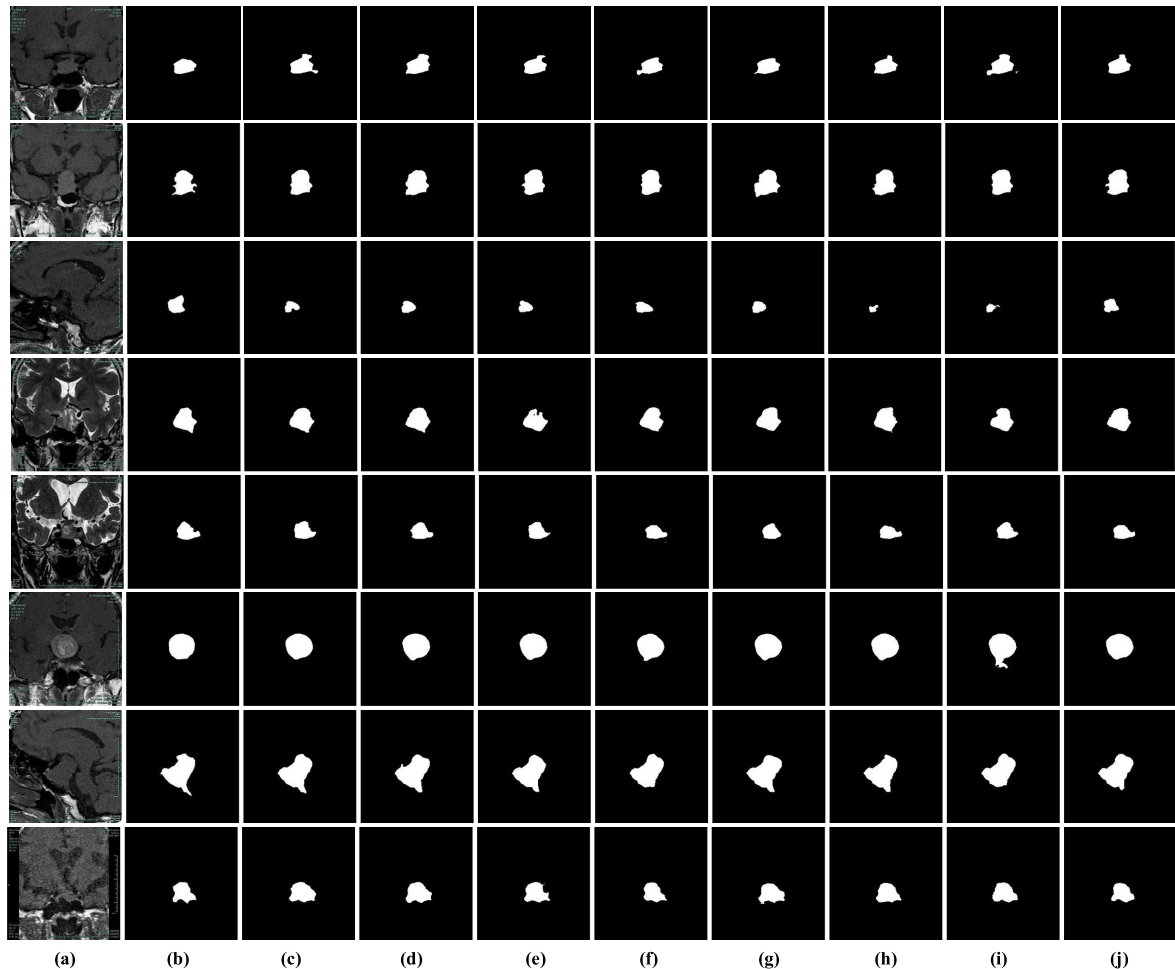
FIGURE 8. Ablation studies of MSR-Net on the original pituitary adenoma dataset. (a-b) are the original images with their corresponding labels. (c-j) are results of Baseline, Baseline+CA, Baseline+SA, Baseline+MSR, Baseline+CA+SA, Baseline+CA+MSR, Baseline+SA+MSR and Baseline+CA+SA+MSR.

was conducted as the baseline network. Among them, Baseline+CA, Baseline+SA, and Baseline+MSR indicate that CA, SA, and MSR modules are added to the U-Net backbone network to provide auxiliary monitoring information; Baseline+CA+SA, Baseline+CA+MSR and Baseline+SA+MSR indicate that MSR, SA, and CA modules are removed from the MSR-Net network; Baseline+CA+SA+MSR indicates the final MSR-Net network structure. The ablation comparative analysis results of each module combination in pituitary adenoma dataset are

shown in Table 4. Compared with Baseline, Baseline+CA, Baseline+SA, and Baseline+MSR were all improved in the indicators of specificity, IoU, Mcc and Dice, and the MSR-Net network based on Baseline+CA+SA+MSR showed the greatest improvement. As can be seen from Figure 8, in the image with large feature difference, low contrast and fuzzy boundary in the target region, more accurate detailed boundary of the target region can be obtained through the fusion of channel attention, spatial attention and multi-scale residual features. Therefore, the results

of the ablation experiment underscore the importance of each module in the MSR-Net framework, which validated the effectiveness of improving segmentation accuracy and robustness in the diagnosis of pituitary adenoma.

E. COMPARISON WITH OTHER ALGORITHMS

To thoroughly assess the advanced of MSR-Net, the performance of other methods was quantitatively compared. In the same experimental environment, U-Net [6], BCDU-Net [31], OD-Net [32], CLCI-Net [33], DUDA-Net [34], FF-UNet [35], TAUnet [36], PDC-Net [37], MPSU-Net [38], and PMENet [39] were trained using the original pituitary adenoma dataset, and corresponding segmentation results were obtained. The results of quantitative evaluation are

meticulously documented in Table 5, in which the best results are in black body. As shown in the table, the MSR-Net achieved optimal segmentation results for specificity, IoU, Mcc and Dice, whose values were 99.74%, 80.87%, 89.12%, and 89.34%. Compared with the second ranked segmentation method, the segmentation results increased by 0.02%, 1.96%, 1.18% and 1.18%, respectively. After comprehensive and comparative analysis, our algorithm achieved the best performance in the lesion segmentation task on the original pituitary adenoma dataset.

To further analyze the lesion segmentation effect, visual segmentation and comparison results were performed on the pituitary adenoma dataset, as depicted in Figure 9. This examination revealed a myriad of complexities inherent in the lesion structures within the images, including complex

TABLE 5. Comparative experiments of different networks on the original pituitary adenoma dataset.

Method	Specificity	IoU	Mcc	Dice
U-Net [6]	0.9958	0.7491	0.8536	0.8558
BCDU-Net [31]	0.9963	0.7815	0.8739	0.8762
OD-Net [32]	0.9952	0.7217	0.8348	0.8369
CLCI-Net [33]	0.9960	0.7374	0.8449	0.8477
DUDA-Net [34]	0.9962	0.7432	0.8482	0.8511
FF-UNet [35]	0.9954	0.7440	0.8505	0.8523
TAUnet [36]	0.9954	0.7375	0.8459	0.8478
PDC-Net [37]	0.9970	0.7891	0.8794	0.8816
MPSU-Net [38]	0.9968	0.7714	0.8677	0.8703
PMENet [39]	0.9972	0.7867	0.8776	0.8800
MSR-Net	**0.9974**	**0.8087**	**0.8912**	**0.8934**

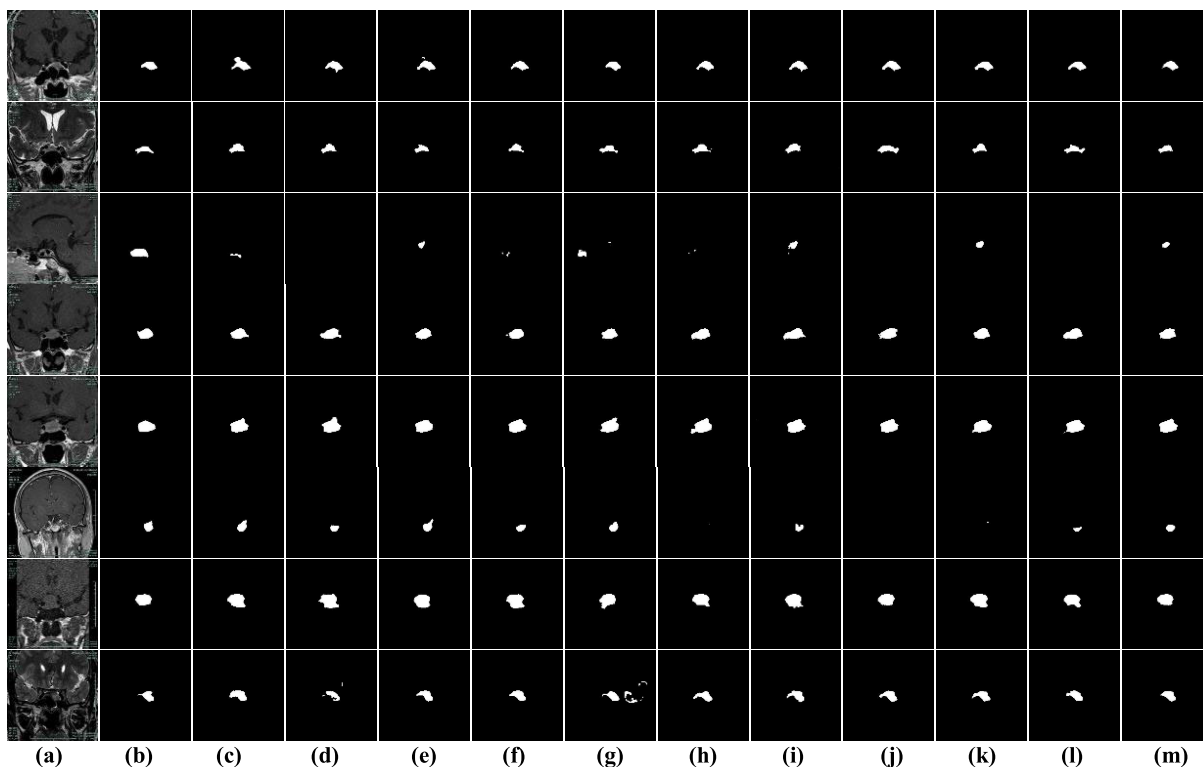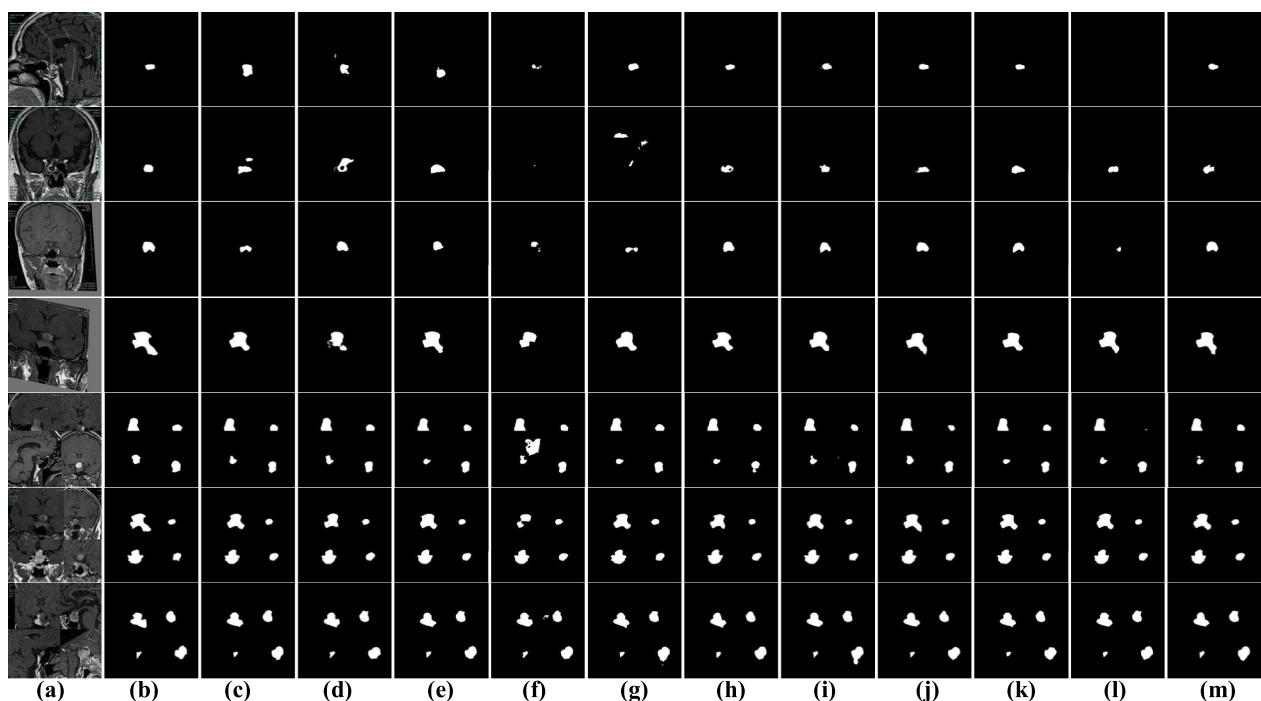| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) | (i) | (j) | (k) | (l) | (m) |

FIGURE 9. Visualization comparison results on the original pituitary adenoma dataset. (a) Original images; (b) corresponding labels; (c-m) are results of U-Net, BCDU-Net, OD-Net, CLCI-Net, DUDA-Net, FF-UNet, TAUnet, PDC-Net, MPSU-Net, PMENet and MSR-Net.

TABLE 6. Comparative experiments of different networks on the enhanced pituitary adenoma dataset.

Method	Specificity	IoU	Mcc	Dice
U-Net [6]	0.9956	0.7902	0.8790	0.8819
BCDU-Net [31]	0.9951	0.7659	0.8630	0.8664
OD-Net [32]	0.9838	0.5732	0.7318	0.7254
CLCI-Net [33]	0.9965	0.7969	0.8830	0.8862
DUDA-Net [34]	0.9956	0.7659	0.8626	0.8662
FF-UNet [35]	0.9964	0.7926	0.8803	0.8836
TAUnet [36]	0.9910	0.6905	0.8145	0.8153
PDC-Net [37]	0.9967	0.8083	0.8903	0.8856
MPSU-Net [38]	0.9958	0.7885	0.8778	0.8811
PMENet [39]	0.9962	0.7958	0.8824	0.8934
MSR-Net	**0.9969**	**0.8096**	**0.8912**	**0.8942**

FIGURE 10. Visualization comparison results on the enhanced pituitary adenoma dataset. (a) Original images; (b) corresponding labels; (c-m) are results of U-Net, BCDU-Net, OD-Net, CLCI-Net, DUDA-Net, FF-UNet, TAUnet, PDC-Net, MPSU-Net, PMENet and MSR-Net.

boundaries, irregular shape, and uneven strength of surrounding tissues. According to the segmentation results, BDU-Net, CLCI-Net, DUDA-Net, FF-UNet, TAUnet, PDC-Net, and PMENet have some mis-segmentation regions, while other algorithms all lose some contour details at the lesion boundaries to varying degrees. In summary, compared with other networks, the MSR-Net can retain more detailed information inside the focal area and at complex boundaries without any post-processing. At the same time, it can effectively avoid the interference of the uneven strength of the surrounding tissue, and the segmentation result is closer to the corresponding labels image.

To showcase the stability and adaptability of MSR-Net, we conducted a meticulous comparative experiment using the enhanced pituitary adenoma dataset. As highlighted in Table 6, there was a noteworthy reduction in performance indicators for OD-Net and TAUnet, indicating potential limitations in accurately segmenting pituitary adenoma lesions with complex shapes and deformations. MPSU-Net focuses on multi-path supervision to refine feature extraction, showing improvements over U-Net but not surpassing the top-performing models. With its attention mechanisms and enhanced learning modules, PMENet achieves high performance, particularly in the Dice coefficient, indicating better overlap with ground truth labels. Conversely, the performance metrics of other models remained relatively stable. However, our MSR-Net integrates multi-scale residual blocks and attention mechanisms to capture both local and global features effectively, which achieve the highest IoU, MCC, and Dice coefficient scores on the enhanced pituitary adenoma dataset. The exceptional segmentation accuracy demonstrated by MSR-Net underscores its potential significance for enabling precise lesion segmentation and boundary delineation in clinical settings. Furthermore, the visualization results of the

above methods, as shown in Figure 10, provide compelling evidence for the superior performance of MSR-Net, further enhancing its confidence and applicability in medical image analysis.

IV. CONCLUSION

According to the complex characteristics of pituitary adenoma, such as variable shape and location, a new segmentation framework based on multi-scale residual, channel attention and spatial attention was developed to further enhance the ability of feature expression. The relevant conclusions are summarized as: firstly, a multi-scale residual module is designed to enhance the capability to extract channel features at different scales. Secondly, two paths of channel attention and spatial attention are constructed to obtain the response weight and global dependence of each channel to key information. Experiments show that MSR-Net achieves 99.74% of specificity, 80.87% of IoU, 89.12% of Mcc, 89.34% of Dice on the original pituitary adenoma dataset and 99.67% of specificity, 80.96% of IoU, 89.12% of Mcc, 89.42% of Dice. These results significantly outperform other state-of-the-art algorithms, underscoring the effectiveness and superiority of the proposed method in the segmentation of pituitary adenoma.

Although MSR-Net performs well in the segmentation of pituitary adenomas, its efficiency and adaptability still need to be explored and improved. In future work, further refinement of the framework will be explored, along with its application to other types of medical imaging tasks. Additionally, the integration of this framework with other diagnostic tools and its potential for clinical adoption will be studied to improve the overall accuracy and efficiency of the diagnosis and treatment of pituitary adenoma.

REFERENCES

[1] D. Morita, S. Mazen, S. Tsujiko, Y. Otake, Y. Sato, and T. Numajiri, "Deep-learning-based automatic facial bone segmentation using a two-dimensional U-Net," *Int. J. Oral Maxillofacial Surg.*, vol. 52, no. 7, pp. 787–792, Jul. 2023.

[2] H. Dang, M. Li, X. Tao, G. Zhang, and X. Qi, "LVSegNet: A novel deep learning-based framework for left ventricle automatic segmentation using magnetic resonance imaging," *Comput. Commun.*, vol. 208, pp. 124–135, Aug. 2023.

[3] W. Zou, X. Qi, W. Zhou, M. Sun, Z. Sun, and C. Shan, "Graph flow: Cross-layer graph flow distillation for dual efficient medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 4, pp. 1159–1171, Apr. 2023.

[4] X. Shu, Y. Yang, J. Liu, X. Chang, and B. Wu, "ALVLS: Adaptive local variances-based levelset framework for medical images segmentation," *Pattern Recognit.*, vol. 136, Apr. 2023, Art. no. 109257.

[5] 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.

[6] 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.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[7] P. Tang, P. Yang, D. Nie, X. Wu, J. Zhou, and Y. Wang, "Unified medical image segmentation by learning from uncertainty in an end-to-end manner," *Knowledge-Based Syst.*, vol. 241, Apr. 2022, Art. no. 108215.

[8] J. Sun, C. Li, Z. Lu, M. He, T. Zhao, X. Li, L. Gao, K. Xie, T. Lin, J. Sui, Q. Xi, F. Zhang, and X. Ni, "TNSNet: Thyroid nodule segmentation in ultrasound imaging using soft shape supervision," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, Art. no. 106600.

[9] H. Zhang, X. Zhong, Z. Li, Y. Chen, Z. Zhu, J. Lv, C. Li, Y. Zhou, and G. Li, "TiM-Net: Transformer in M-Net for retinal vessel segmentation," *J. Healthcare Eng.*, vol. 2022, pp. 1–17, Jul. 2022.

[10] Z. Fu, J. Zhang, R. Luo, Y. Sun, D. Deng, and L. Xia, "TF-UNet: An automatic cardiac MRI image segmentation method," *Math. Biosci. Eng.*, vol. 19, no. 5, pp. 5207–5222, 2022.

[11] Y. Zhang, J. Chu, L. Leng, and J. Miao, "Mask-refined R-CNN: A network for refining object details in instance segmentation," *Sensors*, vol. 20, no. 4, p. 1010, Feb. 2020.

[12] C. Peng, M. Zhu, H. Ren, and M. Emam, "Small object detection method based on weighted feature fusion and CSMA attention module," *Electronics*, vol. 11, no. 16, p. 2546, Aug. 2022.

[13] J. Lu, H. Ren, M. Shi, C. Cui, S. Zhang, M. Emam, and L. Li, "A novel hybridoma cell segmentation method based on multi-scale feature fusion and dual attention network," *Electronics*, vol. 12, no. 4, p. 979, Feb. 2023.

[14] L. Yuan, J. Song, and Y. Fan, "MCNMF-UNet: A mixture conv-MLP network with multi-scale features fusion UNet for medical image segmentation," *PeerJ Comput. Sci.*, vol. 10, p. e1798, Jan. 2024.

[15] M. U. Saeed, W. Bin, J. Sheng, H. M. Albarakati, and A. Dastgir, "MSFF: An automated multi-scale feature fusion deep learning model for spine fracture segmentation using MRI," *Biomed. Signal Process. Control*, vol. 91, May 2024, Art. no. 105943.

[16] 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.

[17] M. R. Ahmed, A. F. Ashrafi, R. U. Ahmed, S. Shatabda, A. K. M. M. Islam, and S. Islam, "DoubleU-NetPlus: A novel attention and context-guided dual U-Net with multi-scale residual feature fusion network for semantic segmentation of medical images," *Neural Comput. Appl.*, vol. 35, no. 19, pp. 14379–14401, Jul. 2023.

[18] T. J. Yang, Q. Tang, L. Li, and X. H. Bai, "Non-rigid medical image registration using multi-scale residual deep fully convolutional networks," *J. Instrum.*, vol. 16, no. 3, Mar. 2021, Art. no. P03005.

[19] M. Sun, P. Li, J. Ren, and Z. Wang, "Attention mechanism enhanced multi-layer edge perception network for deep semantic medical segmentation," *Cognit. Comput.*, vol. 15, no. 1, pp. 348–358, Jan. 2023.

[20] T. Shan, J. Yan, X. Cui, and L. Xie, "DSCA-Net: A depthwise separable convolutional neural network with attention mechanism for medical image segmentation," *Math. Biosci. Eng.*, vol. 20, no. 1, pp. 365–382, 2022.

[21] 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.

[22] 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.

[23] X. Xie, X. Pan, W. Zhang, and J. An, "A context hierarchical integrated network for medical image segmentation," *Comput. Electr. Eng.*, vol. 101, Jul. 2022, Art. no. 108029.

[24] 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.

[25] I. Ahmed, M. Ahmad, and G. Jeon, "A real-time efficient object segmentation system based on U-Net using aerial drone images," *J. Real-Time Image Process.*, vol. 18, no. 5, pp. 1745–1758, Oct. 2021.

[26] C. Huang and M. Wan, "Automated segmentation of brain tumor based on improved U-Net with residual units," *Multimedia Tools Appl.*, vol. 81, no. 9, pp. 12543–12566, Apr. 2022.

[27] J. Li, S. Wang, S. Hu, Y. Sun, Y. Wang, P. Xu, and J. Ye, "Class-aware attention network for infectious keratitis diagnosis using corneal photographs," *Comput. Biol. Med.*, vol. 151, Dec. 2022, Art. no. 106301.

[28] M. Jiang, F. Zhai, and J. Kong, "A novel deep learning model DDU-Net using edge features to enhance brain tumor segmentation on MR images," *Artif. Intell. Med.*, vol. 121, Nov. 2021, Art. no. 102180.

[29] 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.

[30] 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.

[31] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with Densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 406–415.

[32] 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.

[33] H. Yang, W. Huang, K. Qi, C. Li, X. Liu, M. Wang, H. Zheng, and S. Wang, "CLCI-Net: Cross-level fusion and context in-Ference networks for lesion segmentation of chronic stroke," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent*. Cham, Switzerland: Springer, 2019, pp. 266–274.

[34] F. Xie, Z. Huang, Z. Shi, T. Wang, G. Song, B. Wang, and Z. Liu, "DUDA-net: A double U-shaped dilated attention network for automatic infection area segmentation in COVID-19 lung CT images," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 16, no. 9, pp. 1425–1434, Sep. 2021.

[35] A. Iqbal, M. Sharif, M. A. Khan, W. Nisar, and M. Alhaisoni, "FF-UNet: A U-shaped deep convolutional neural network for multimodal biomedical image segmentation," *Cognit. Comput.*, vol. 14, no. 4, pp. 1287–1302, Jul. 2022.

[36] D. Li, Y. Peng, Y. Guo, and J. Sun, "TAUNet: A triple-attention-based multi-modality MRI fusion U-Net for cardiac pathology segmentation," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2489–2505, Jun. 2022.

[37] Q. Zhang, J. Cheng, C. Zhou, X. Jiang, Y. Zhang, J. Zeng, and L. Liu, "PDC-net: Parallel dilated convolutional network with channel attention mechanism for pituitary adenoma segmentation," *Frontiers Physiol.*, vol. 14, Aug. 2023, Art. no. 1259877.

[38] B. Wang, J. Li, C. Dai, W. Zhang, and M. Zhou, "MPSU-net: Quantitative interpretation algorithm for road cracks based on multiscale feature fusion and superimposed U-Net," *Digit. Signal Process.*, vol. 153, Oct. 2024, Art. no. 104598.

[39] B. Wang, C. Dai, J. Li, X. Jiang, J. Zhang, and G. Jia, "PMENet: A parallel UNet based on the fusion of multiple attention mechanisms for road crack segmentation," *Signal, Image Video Process.*, vol. 18, no. S1, pp. 757–769, Aug. 2024.

XIAOLIANG JIANG received the M.S. and Ph.D. degrees in mechanical design from Southwest Jiaotong University, Chengdu, China. He is currently a Professor with Quzhou University. His research interests include machine vision and image recognition.

XIUQING HUANG received the bachelor's degree in medicine from Wenzhou Medical University. He is currently the Chief Physician of The Quzhou Hospital, Wenzhou Medical University. His research interests include diagnosis and imaging analysis of neurological diseases.

QILE ZHANG received the master's degree in medicine from Kunming Medical University. She is currently the Deputy Chief Physician of The Quzhou Hospital, Wenzhou Medical University. Her research interests include diagnosis and imaging analysis of neurological diseases.

CHUN ZHOU received the master's degree in medicine from Zhejiang Chinese Medical University. She is currently the Deputy Chief Physician of The Quzhou Hospital, Wenzhou Medical University. Her research interests include the diagnosis and treatment of neurological diseases.

• • •