



# Multi-scale information residual network: Deep residual network of prostate cancer segmentation based on multi scale information guidance



Xinyi Chen<sup>a</sup>, Xiang Liu<sup>a,\*</sup>, Yunjie Yu<sup>b,\*</sup>, Yunyu Shi<sup>a</sup>, Yuke Wu<sup>a</sup>, Zhenglei Wang<sup>a</sup>, Shuhong Wang<sup>c</sup>

<sup>a</sup> School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

<sup>b</sup> Shanghai Electric Power Hospital, Shanghai 201620, China

<sup>c</sup> Department of Molecular and Cellular Biology and Center for Brain Science, Harvard University, Cambridge, MA 02138, USA

## ARTICLE INFO

**Keywords:**  
Prostate Cancer  
MSR-Net  
DeepLearning  
Segmentation  
MIGF model

## ABSTRACT

Magnetic resonance imaging (MRI) is a key tool in prostate cancer screening and diagnosis, with automatic segmentation of the cancer crucial for accurate staging and treatment. Nevertheless, the accurate segmentation of prostate cancer remains a challenging subject. In order to address this challenge, a two-stage segmentation method is employed. In the initial stage, the prostate tissue is delineated as the region of interest. Subsequently, in the second stage, the precise segmentation of prostate cancer is achieved on the aforementioned region of interest. In order to accurately segment the region of interest, we propose MSR-Net (Multi-scale information residual network), which employs an MSR-block, designed based on MLKA convolution, to extract multi-scale information, combines the group attention mechanism to enhance the multi-scale features, and uses the improved CGA feature fusion module to fuse deep and shallow features. The feature map is then upsampled using DySample. The experiments conducted on the Prostataex dataset for the segmentation of prostate cancer were based on the Dice metric. The results demonstrate an improvement of 5.2% (60.5% vs. 55.3%) in comparison to the second-best performing segmentation network (Unet). The application of the two-stage segmentation method has a considerable impact, with a 10.4% improvement (45.3% vs 55.7%) on the baseline when two-stage segmentation is employed for prostate cancer in comparison to direct segmentation of prostate cancer. Furthermore, the network was trained and tested on the prostate segmentation and lung nodule segmentation datasets, achieving the highest dice scores of 0.937 and 0.764, respectively.

## 1. Introduction

Prostate cancer (PCA) is the second most prevalent form of cancer and the fifth leading cause of cancer-related mortality among men. The incidence rates of prostate cancer (PCA) are 37.5 per 100,000 in developed countries and 11.3 per 100,000 in developing countries, while mortality rates are 8.1 per 100,000 in developed countries and 5.9 per 100,000 in developing countries [1]. In comparison to other forms of cancer, such as lung and liver cancer, there is a notable lack of awareness regarding prostate cancer screening. Additionally, the academic research landscape surrounding prostate cancer is relatively underdeveloped, and diagnostic tools are scarce. Consequently, patients frequently present with advanced-stage prostate cancer. At present, the diagnostic tools available for prostate cancer include puncture biopsy

and medical imaging techniques such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound. Puncture biopsy of the prostate represents the gold standard for the diagnosis of prostate cancer. Clinically, a significant proportion of prostate cancers develop slowly in the early stages. Consequently, the diagnosis of whether a patient is suffering from prostate cancer must be confirmed by puncture biopsy of the prostate. Nevertheless, as the definitive test for prostate cancer, puncture biopsy is an invasive procedure that carries the risk of bleeding and pain [2]. Consequently, a biopsy should only be conducted when there is a high probability that the patient has prostate cancer. It is therefore recommended that prostate cancer be screened using medical imaging methods prior to performing a puncture biopsy for prostate cancer.

The use of medical imaging in the diagnosis of prostate cancer is a

\* Corresponding authors.

E-mail addresses: [c2257873708@163.com](mailto:c2257873708@163.com) (X. Chen), [xliu@sues.edu.cn](mailto:xliu@sues.edu.cn) (X. Liu), [yuyunjie186@163.com](mailto:yuyunjie186@163.com) (Y. Yu), [yunyushi@sues.edu.cn](mailto:yunyushi@sues.edu.cn) (Y. Shi), [M325122219@sues.edu.cn](mailto:M325122219@sues.edu.cn) (Y. Wu), [hanqi\\_willis@163.com](mailto:hanqi_willis@163.com) (Z. Wang), [wangsh@fas.harvard.edu](mailto:wangsh@fas.harvard.edu) (S. Wang).

crucial element in the early detection, auxiliary diagnosis, risk stratification, guidance of treatment and efficacy assessment of this disease [1,3]. The accurate extraction of cancerous areas from images can facilitate a more comprehensive understanding of the lesions and pathological changes, thereby enhancing the accuracy of prostate cancer diagnosis and staging. Magnetic resonance imaging (MRI) represents a principal imaging modality for the diagnosis of prostate cancer. Prostate cancer lesions on MR images present with a variety of shapes and sizes, and are often subtle and challenging to discern. Consequently, the accurate segmentation of prostate cancer lesions remains a significant area of focus within the field of medical image processing. Presently, the segmentation of prostate cancer is predominantly conducted manually. However, manual segmentation of prostate cancer is challenging, necessitating significant effort and expertise. It also requires doctors to possess a high level of medical literacy.

In order to enhance the efficacy of segmentation and diminish the workload of radiologists while augmenting the precision of segmentation, an increasing number of researchers are directing their attention towards the automatic segmentation of the prostate and prostate cancer. The advent of deep learning and the growing prevalence of medical imaging have led to a surge in the use of convolutional neural networks (CNNs) for medical image segmentation. A variety of convolutional neural network models have been employed in prostate segmentation research, including U-Net [4,5], FCN [6], and others, which have demonstrated superior performance in prostate segmentation.

Despite the existence of a considerable number of studies on deep learning for prostate segmentation, which yield relatively positive results, there is a paucity of research focusing on prostate cancer segmentation. Furthermore, there is considerable scope for improvement in these studies. In order to achieve efficient automatic prostate cancer segmentation, we propose an attention-based hybrid convolutional neural network. A hybrid convolutional module is designed based on MLKA [7] as the encoder and decoder of the network. This introduces MLKA convolution, which enables the network to adaptively capture long-range dependency and local granularity information. The module is also based on the residual connection method, which enables the network to better fuse the feature information of different paths. This allows the model to combine the information of different feature levels, thereby learning richer representations and improving the robustness of the model. The network incorporates the SpatialGroupEnhance(SGE) [8] module at jump connections to enhance the representation of features in the spatial dimension. This enables the model to more effectively identify and focus on the most crucial aspects of these multi-level features, preventing the accumulation of redundant information and the interference of irrelevant features. Furthermore, the CGAFusion [9] feature fusion curvature has been enhanced to enable the network to more effectively fuse shallow and deep features. In the upsampling stage, DySample [10] dynamic upsampling, Venus and dynamic cheap sampling are employed to adaptively adjust the sampling position, thereby facilitating the capture of image details and structure. The contributions of this paper are as follows:

for the problem of difficult segmentation of prostate cancer, we propose MSR-Net for prostate cancer segmentation task, and at the same time, we apply this network to prostate region segmentation task and lung nodule lesion region segmentation task, and in these segmentation tasks, compared with other methods, this network achieves the best results.

For the problem of large scale variability of prostate cancer lesions, we design a multi-scale feature extraction enhancement network (MSR-Block), which extracts multi-scale information from lesion images and fuses the features with residual connection to accurately extract the size and location information of prostate cancer, and improves the accuracy of segmentation.

Based on CGAFusion, the MIGF feature fusion module is proposed to fuse the shallow and deep features of the network to make full use of

the image features, which pays more attention to the connection between the global and local features, so that it can better focus on the key regions and important features in the complex scene.

Applying two-stage segmentation method to segment prostate cancer; the segmentation accuracy can be effectively improved by this method.

## 2. Related work

### 2.1. Research on medical image segmentation

The advancement of medical imaging and deep learning has led to the emergence of numerous techniques for processing medical images, including diagnosis, denoising, and segmentation. Among these, the segmentation of tissue and lesion regions represents a pivotal area of focus in medical image processing.

The Fully Convolutional Networks (FCN) framework, proposed by Jonathan Long et al. in 2015, represents a pioneering contribution to the field of deep learning for semantic segmentation. It is one of the earliest models applied to medical images, with applications including segmentation of skin diseases [11], lumbar spine segmentation [12], thyroid nodule segmentation [13], and others. On the basis of FCN, Olaf Ronneberger et al. [14] proposed the Unet network, which has become the most widely used baseline in medical image segmentation. The U-net incorporates a jump connection structure, which enables the preservation of contextual information and details. Furthermore, the model employs a multitude of data enhancement techniques to enhance its robustness and generalisation capabilities. It has demonstrated efficacy in medical image segmentation, particularly in accurately delineating regions such as tumours and brain structures. Moreover, numerous derivative networks derived from U-Net have also demonstrated efficacy in medical image segmentation. Examples include Swin-Unet [15], Unet++ [16], and NNUNet [17], which have exhibited remarkable performance and have been extensively utilized in the domain of medical image segmentation research.

### 2.2. Research on the prostate gland segmentation

Given the extensive use of deep learning in medical imaging, numerous studies have been conducted on the segmentation of the prostate. One such study, proposed by Tian et al., [18] introduced PSNet, a method for prostate segmentation based on the enhancement of a full convolutional neural network. This approach involves predicting the probability that each voxel belongs to the prostate or non-prostate, thereby facilitating the segmentation of the prostate. Fan Song et al. [19] proposed PSP-Net for prostate segmentation, utilising a residual structure to construct an effective a priori feature extraction network and formulate an optimal policy rate, and inputting 2D images obtained from slices of 3D prostate maps for training and testing. Renato et al. [20] evaluated the performance of E-Net, U-Net, and ERF-Net for prostate segmentation. Bardis et al. [21] cascaded three U-Nets on a UNet-based network. U-NetA was used for localisation, with the localised prostate region input into U-NetB and U-NetC. U-NetB was used to segment the prostate region, with the segmented prostate region inputted into U-NetC. U-NetC was used to segment the transition and peripheral regions.

### 2.3. Research on prostate cancer segmentation

Despite the success of prostate segmentation, there is still a large gap in prostate cancer segmentation, which is difficult to observe on MR because of the variable size and shape of prostate cancer, which poses a major difficulty in lesion segmentation. Sherif Mehralivand et al. [22] designed a cascaded deep learning network to detect and segment prostate cancer lesions using cascaded UNet and residual networks. The median DSC for lesion segmentation was 0.359. Yue Lin et al. [23]

evaluated a two-parameter MRI AI algorithm for the detection and segmentation of intraprostatic lesions and compared its performance with radiologist's readings and biopsy results. Based on the deep learning approach of Sherif Mehralivand et al. [22] the performance of the AI algorithm for detecting suspicious cancerous lesions on two-parameter MRI scans was found to be satisfactory. suspicious lesions of cancer with a performance comparable to that of an experienced radiologist. However, its lesion segmentation DSC was 0.29.

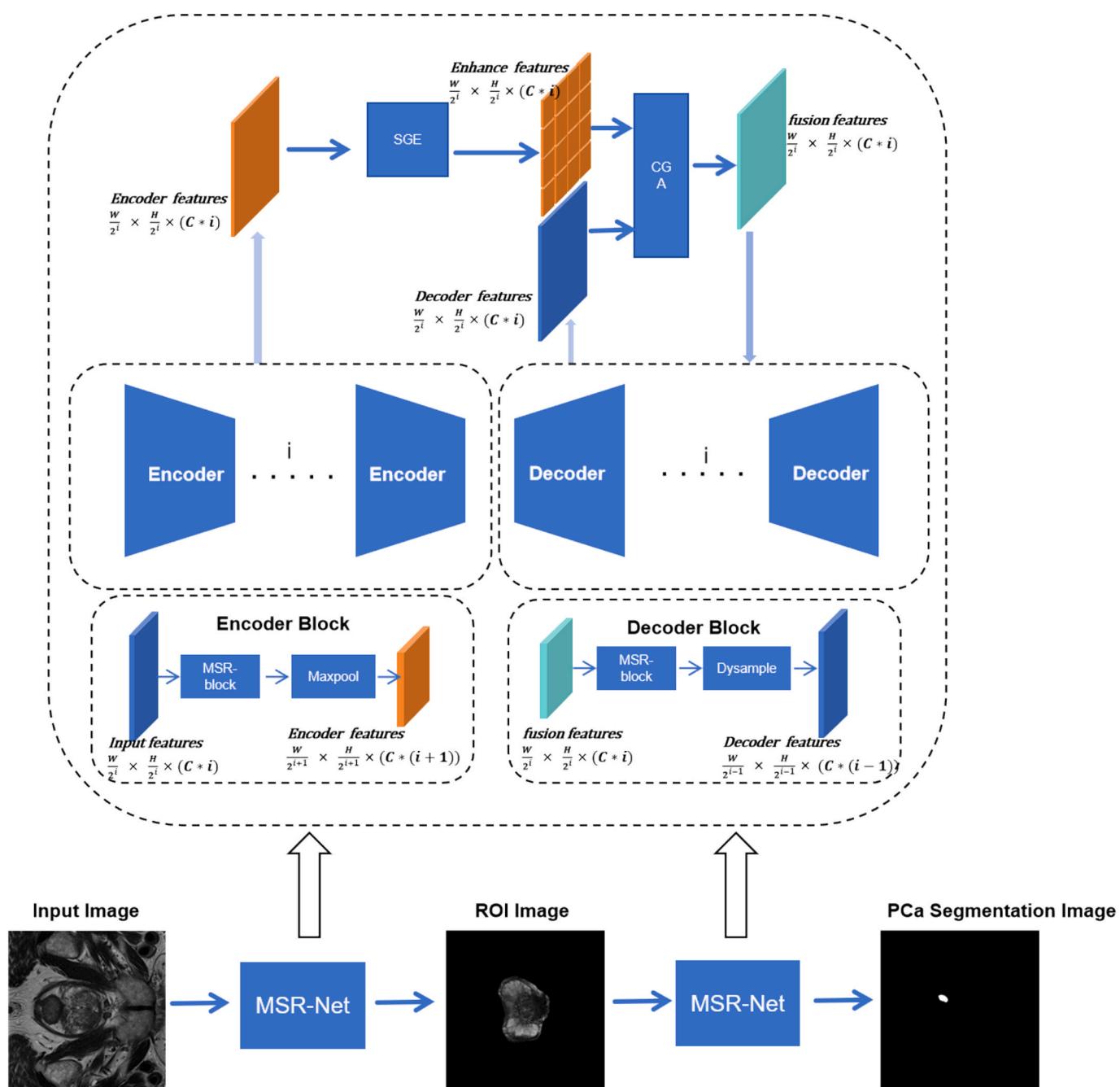
#### 2.4. Conclusion

In the field of prostate cancer segmentation, despite the significant progress made by deep learning-based methods, these approaches still

have some limitations and challenges. Deep learning models in the field of prostate cancer segmentation have shown significant advantages in terms of accuracy and automation. These models are capable of effectively processing complex medical images and learning directly from raw data, eliminating the need for manual feature engineering.

However, these models often struggle with detecting minor and less apparent lesions, making them unsuitable for early-stage prostate cancer screening. Therefore, this paper proposes a two-stage segmentation method based on MSR-Net for prostate cancer segmentation. This method effectively segments prostate cancer lesions and possesses unique advantages in capturing the nuances of prostate cancer.

In the field of prostate cancer segmentation, existing models often face challenges in accurately identifying minor or less obvious lesions,



**Fig. 1.** The workflow of the two-stage segmentation of prostate cancer. (All abbreviations for this article are explained here:MSR-Net:Multi-scale information residual network ; MR:Magnetic Resonance;ROI: region of interest;CGA: Content-guided Attention;SGE: Spatial Group Enhance ; PCa:Prostate Cancer; GCIE module: Multi-scale information-guided fusion module).

posing significant difficulties for early-stage prostate cancer screening. In light of this, this paper introduces a two-stage segmentation approach based on MSR-Net, designed to precisely segment prostate cancer lesions. This method not only effectively locates and segments tumor regions but also excels in capturing the details of prostate cancer, promising enhanced detection capabilities for subtle lesions. Consequently, it supports earlier and more accurate diagnosis and treatment planning for prostate cancer. The introduction of this method aims to address the deficiencies of current technologies, providing robust support for the early detection and treatment of prostate cancer.

### 3. Methods

#### 3.1. Two-stage prostate cancer segmentation methods

Magnetic resonance imaging (MRI) is a diagnostic tool used to examine the prostate gland in patients with suspected prostate cancer. The images produced by MRI are cross-sections of the pelvic region, which include various organs such as the bladder, seminal vesicles, and the prostate. Given that prostate cancer is typically a lesion of the prostate gland, the gland itself is often a relatively minor component of the MR image. Consequently, direct segmentation of prostate cancer on MR maps in cross-section of the pelvic region would be an inadequate approach. In light of the aforementioned circumstances, we elected to implement a two-stage segmentation methodology for the purpose of delineating prostate cancer. As illustrated in Fig. 1, the prostate tissue was initially delineated through network training, and subsequently employed as the region of interest to facilitate the precise segmentation of the prostate cancer lesion area. This approach allows for the effective exclusion of interference from other organs, thereby enhancing the accuracy of the segmentation process. (C is the feature dimension after the first encoder).

#### 3.2. MSR-Net network architecture

Prostate cancer presents with a range of variable sizes and locations, blurred boundaries, and difficulties in observation. Consequently, the segmentation of prostate cancer remains a significant challenge. In order to address the aforementioned characteristics of prostate cancer, we have designed a network architecture, MSR-Net, which comprises a number of key components. These include an MSR-Block-based encoder and decoder, an SGE (Spatial Group Enhancement) feature enhancement module for hopping connections, an MIGF module (Multi-scale information-guided fusion module) and a DySample upsampling module.

The MSR-Block is a feature extraction module that has been designed based on the MLKA convolution and residual structure. The MLKA convolution is based on the LKA idea and introduces a group-by-group multiscale mechanism, whereby the input features are grouped uniformly along the channel direction. Subsequently, kernel decomposition of different sizes is applied to each group in order to obtain features with global and local interpretations. Finally, the output of the grouped features is multiplied element-by-element with the output of the kernel decomposition through the gated coalescence operation, which serves to remove artefacts. Subsequently, the output of DWConv on the grouped features and the output of the kernel decomposition are multiplied element-by-element to remove the artefacts through the gated aggregation operation. Following the MLKA convolution operation, the multiscale information fusion feature can be obtained, which offers a local-global information interpretation and is effective in addressing the issue of prostate cancer with variable size and location. Nevertheless, the concurrent utilisation of multilevel MLKA convolution in conjunction with ordinary convolution results in an elevated complexity of the network and a diminished capacity for network fitting. To address this issue, we propose the introduction of the residual structure, which ensures the smooth propagation of gradients and addresses the common

problem of gradient vanishing and degradation in deep networks. Furthermore, the combination of the residual structure with MLKA convolution allows the MLKA module to fully utilise its capabilities in handling multi-scale information, while the residual structure ensures the delivery of global features. The combination of these two approaches allows the network to achieve a balance between global semantic information and local detailed features. In the context of prostate cancer, where the scale of the lesion region can vary significantly, the integration of MLKA and residual structure enables the effective capture of features at diverse scales, thereby enhancing the identification and characterisation of the lesion region. The pixel attention mechanism is incorporated into the output to facilitate pixel-level weighting calculations, thereby enhancing the capture of detailed information, addressing the issue of prostate cancer edge blurring, and enhancing the network's ability to perform segmentation on edge details.

The MIGF module is based on the CGA (Content-guided Attention) module, a global information-guided deep and shallow information fusion module based on the idea of the CGA-based Mixup Fusion Scheme. The GCIE module has been designed based on global feature-guided enhancement of local information features, with the objective of assisting the model in capturing the similarity between each local region and the overall image, while also enhancing the contextual relationship of the fused features. To address the issue of significant variations in prostate cancer lesion regions, we developed the GCIE module, which is based on global feature guidance to enhance the local information feature enhancement. This is designed to assist the model in capturing the similarity between each local region and the overall image, as well as enhancing the contextual relationship of the fused features. In order to more accurately determine the location and size of prostate cancer, we have adopted a feature splicing method based on the CGA-based Mixup Fusion Scheme. This is followed by the use of PWconv to adjust the output channel, which allows us to retain more original features and enhance the model's expressive ability. Furthermore, the CGA module generates the attention weights of the input features, allowing the network to accurately fuse the shallow features with the deep features. This approach effectively addresses the issue of large-scale differences in the lesion area of prostate cancer.

In addressing the challenge of prostate cancer with indistinct boundaries, we propose the incorporation of the Spatial Group-wise Enhance Attention module at jump junctions. This module enables the adjustment of the relative importance of each sub-feature by generating an attention factor for each spatial location within each semantic group. This approach allows for the autonomous enhancement of the expression of each group's learning, thereby enhancing the CNN representation of the feature map. The spatial enhancement of features by grouping allows the network to be more sensitive to local changes. The utilisation of jump junctions to more effectively combine low-level features from the encoder and high-level features from the decoder enhances the fusion of features and further augments the discriminative capacity of these high-resolution features, thereby facilitating the network's ability to recover details with greater precision. In comparison to the conventional up-sampling techniques (transposed convolution, bilinear interpolation, and so forth), we propose the dynamic up-sampling curation rate DySample, a rapid, effective, and adaptable dynamic up-sampler. This method dynamically modifies the sampling operation in accordance with the content of the feature map, thereby concentrating on the details of the recovered image to achieve a more refined segmentation result and a more detailed segmentation of the leading cancer edge.

Following the input of the image of the prostate cancer lesion, the encoder initially extracts the lesion's deep features. These are then enhanced by the Spatial Group Enhance module, thereby increasing their expressiveness in the spatial dimension. Subsequently, the enhanced features are combined with those of the decoder, and the fusion of shallow and deep features is achieved through the use of the MIGF module. Subsequently, the fused features are up-sampled by the DySample module, and finally, an accurate segmentation result is

produced through a convolutional layer. The network is illustrated in Fig. 2:

Input image  $I \in R^{3 \times H \times W}$ , after the first encoder  $f_{encoder}$  the output  $F_1 \in R^{base\_C \times \frac{H}{2} \times \frac{W}{2}}$ , after multiple encoder's to the feature  $F_i \in R^{(base\_c^i) \times \frac{H}{2^i} \times \frac{W}{2^i}}$ :

$$F_i = f_{encoder}(F_{i-1}) \quad (1)$$

The network has six layers. The feature  $F_{i\_enhance}$  is obtained by passing the feature through the SGE enhancement  $f_{SGE}$  at the jump junction, and the feature fusion is performed using the improved CGA feature fusion module  $f_{CGA}$ .

$$F_{i\_enhance} = f_{SGE}(F_i) \quad (2)$$

$$F_{6\_decoder} = f_{decoder}(F_6) \quad (3)$$

$$F_{5\_decoder} = f_{decoder}(f_{CGA}(F_{5\_decoder}, F_{6\_decoder})) \quad (4)$$

In a similar manner, this results in the final output feature  $F_{out}$ , which is a function of  $F_{out} \in R^{C \times H \times W}$ , where  $C$  is the number of categories.

$$F_{out} = f_{out\_conv}(f_{CGA}(F_{1\_enhance}, F_{1\_decoder})) \quad (5)$$

$f_{out\_conv}$  is the final output layer and consists of a 1x1 convolution with output channel  $C$ .

### 3.3. MSR-Block

In the encoder and decoder stage, we utilise a multi-scale large-kernel convolution residual module (MSR-block), as illustrated in Fig. 3. This module is founded upon the principles of multi-scale large-kernel attention (MLKA) convolution, which enables the capture of contextual information at varying scales and the focusing of attention on the most salient regions of the image through the use of multiple differently sized convolution kernels and attention mechanisms. This module is based on Multi-scale Large Kernel Attention (MLKA) convolution, which employs a combination of multiple differently sized convolutional kernels and attention mechanisms to capture contextual information at varying scales and to direct focus towards the most salient regions of the image. The feature map, which contains rich multi-scale information and context, is subjected to further refinement through the application of a 3x3 convolution, with the objective of enhancing feature localisation. The recursive application of MLKA in conjunction with a 3x3 convolution facilitates a further refinement of the feature extraction process and enables a recursive enhancement of attention, thereby allowing features to be profoundly enhanced in each layer. Furthermore, a residual structure is incorporated to integrate both deep and shallow information, thereby ensuring that the network retains the original structural information even after multiple convolutions and feature extractions. The features obtained from the residuals are subjected to a pixel-level attention mechanism, which enhances the salient features at each point. Subsequently, the data are subjected to an activation function, namely the generalized exponential linear unit (GELU), and then normalized.

For the feature extraction step as shown in the figure, the input feature  $F_{input} \in R^{in\_channels \times H \times W}$  undergoes a 3x3 convolution  $f_{3 \times 3}(\cdot)$ . The feature extraction is performed while changing the input channels to get the initial feature  $F_1 \in R^{out\_channels \times H \times W}$ , as shown below:

$$F_1 = f_{3 \times 3}(F_{input}) \quad (6)$$

Afterwards, after two MLKA with 3x3 convolution consisting of the MLKA feature extraction module  $f_{mlka+3 \times 3}(\cdot)$  The multiscale (MS) output  $F_{MS} \in R^{out\_channels \times H \times W}$  is obtained:

$$F_{MS} = f_{mlka+3 \times 3}(f_{mlka+3 \times 3}(F_1)) \quad (7)$$

Initial output features (IOT)  $F_{IOT} \in R^{out\_channels \times H \times W}$  are obtained by fusing the initial features  $F_1$  with the multiscale depth features  $F_{MS}$  extracted by the MLKA module after residual concatenation:

$$F_{IOT} = F_1 + F_{MS} \quad (8)$$

The resulting preliminary features are passed through a pointwise convolution  $f_{pw}$  and then through a sigmoid to obtain a per-pixel weight map, which is combined with the preliminary output features FIOT element-wise multiplication  $\odot$  to obtain the final output  $F_{out} \in R^{out\_channels \times H \times W}$ . The final output  $F_{out} \in R^{out\_channels \times H \times W}$  is obtained from the preliminary output features:

$$F_{out} = F_{IOT} \odot \text{sigmod}(f_{pw}(F_{IOT})) \quad (9)$$

### 3.4. Multi-scale information-guided fusion module

In the stage of shallow and deep feature fusion, we design the global context enhancement module (GCIE module) and improve the feature fusion module CGA fusion based on it to get the new feature fusion module MIGF module (multi-scale information-guided fusion module), which makes it more focused on the connection between global and local features. The module includes a global context information enhancement module, a feature weight assignment module and a feature weighted fusion module. The global context information enhancement module summarises the information of the whole spatial dimension by the dot product of each group of features in the feature map and their global pooled features, which helps the model to capture the similarity between each local region and the whole image, so that the network better understands the relationship between the current local features and the whole context, and highlights the important and relevant features; the feature weight assignment module is based on spatial attention, channel attention and pixel attention, and combines with pixel attention to determine the relationship between collocated features and the whole context. The feature weight assignment module is based on spatial attention, channel attention and pixel attention, and combines with pixel attention to determine the relationship between the collocated features and the overall context; the feature weight assignment module is based on spatial attention, channel attention and pixel attention, and combines with pixel attention to determine the relationship between the collocated features and the overall context; the feature weight assignment module combines with pixel attention to perform attention operations on the collocated features, combines local features with global information, weights the features through multiple attention mechanisms, obtains the weights occupied by each feature, and finally performs feature-weighted fusion to accurately and efficiently fuse the deep features with the shallow features, so as to better focus on the key regions and important features in the complex scene.

As shown in Fig. 4, in the global feature enhancement module, we capture the relationship between local features and global features by dot-producing each set of features in space with its global pooled features, as follows, the input feature  $F_a \in R^{C \times H \times W}$ , which is dot-produced  $\odot$  with the original feature after global average pooling gap and then summed with the original feature to get the output  $F_a \in R^{C \times H \times W}$

$$F_a = F_a * \text{gap}(F_a) + F_a \quad (10)$$

The input features are  $F_a, F_b$ , respectively, and the features are subjected to feature splicing  $f_{cat}$  to obtain the intermediate feature  $F_{middle} \in R^{(2^*C) \times H \times W}$ . Meanwhile, the input features  $F_a, F_b$  are  $F_a, F_b \in R^{C \times H \times W}$ , respectively, after the global feature enhancement module  $f_{GCIE}$ .

$$F_a, F_b = f_{GCIE} F_a, F_b \quad (11)$$

$$F_{middle} = f_{cat}(F_a, F_b) \quad (12)$$

The intermediate feature  $F_{middle}$  goes through the CGA,  $f_{CGA}$  to obtain

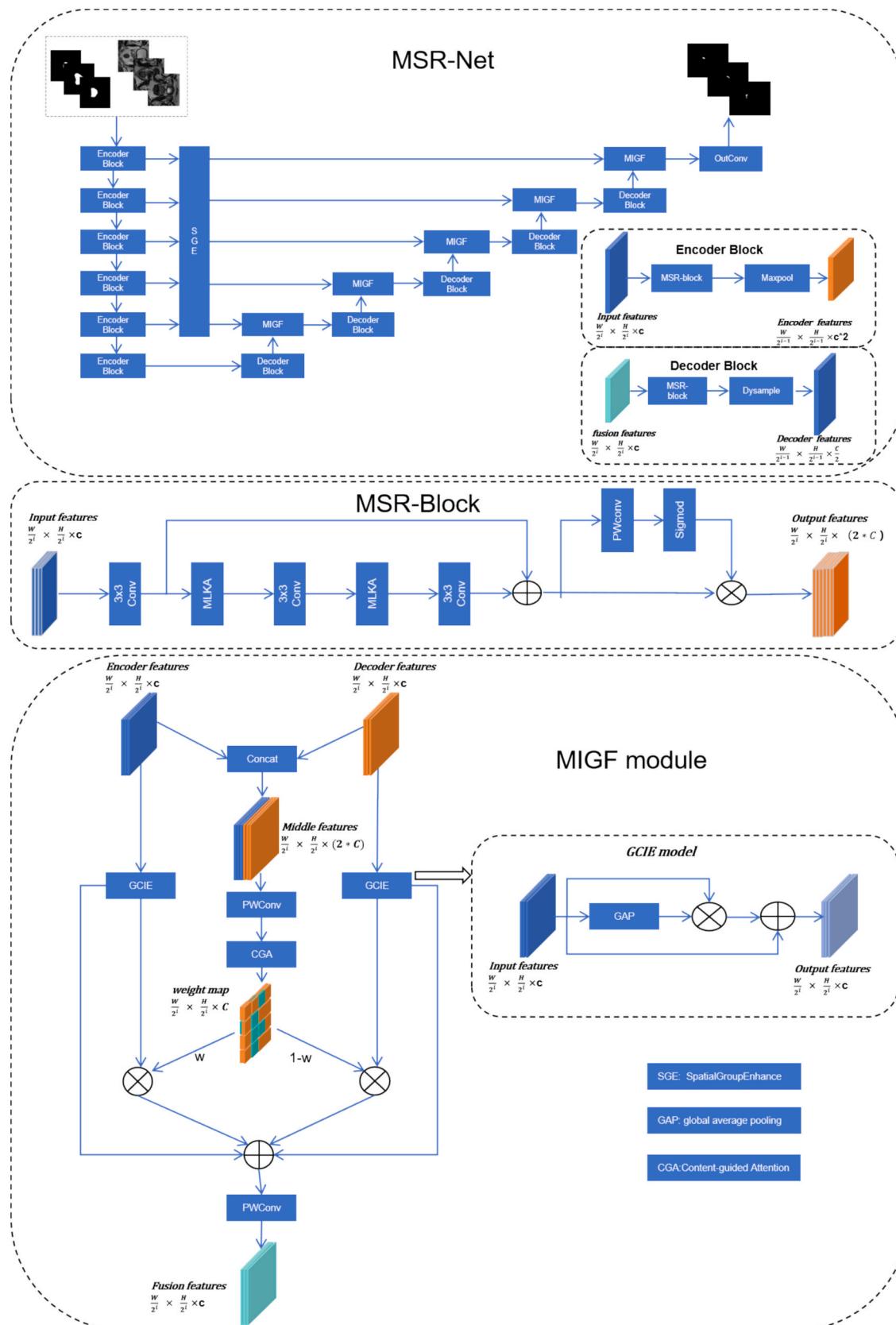


Fig. 2. Overall structure of the Multi-scale information residual network (MSR-Net).

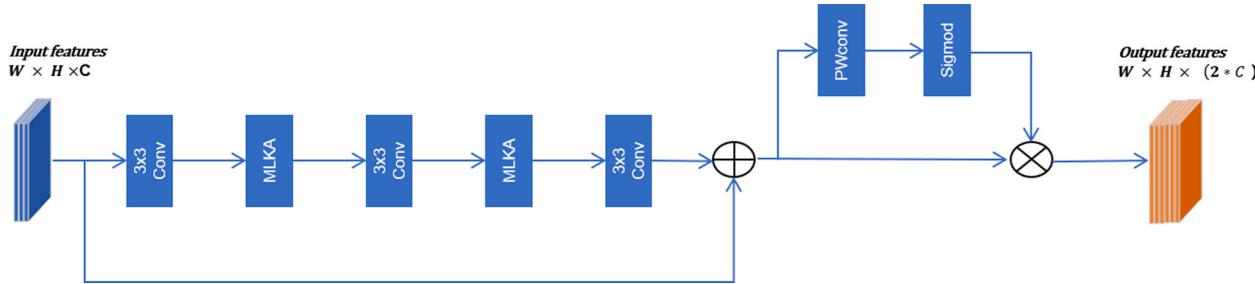


Fig. 3. MSR-block structure.

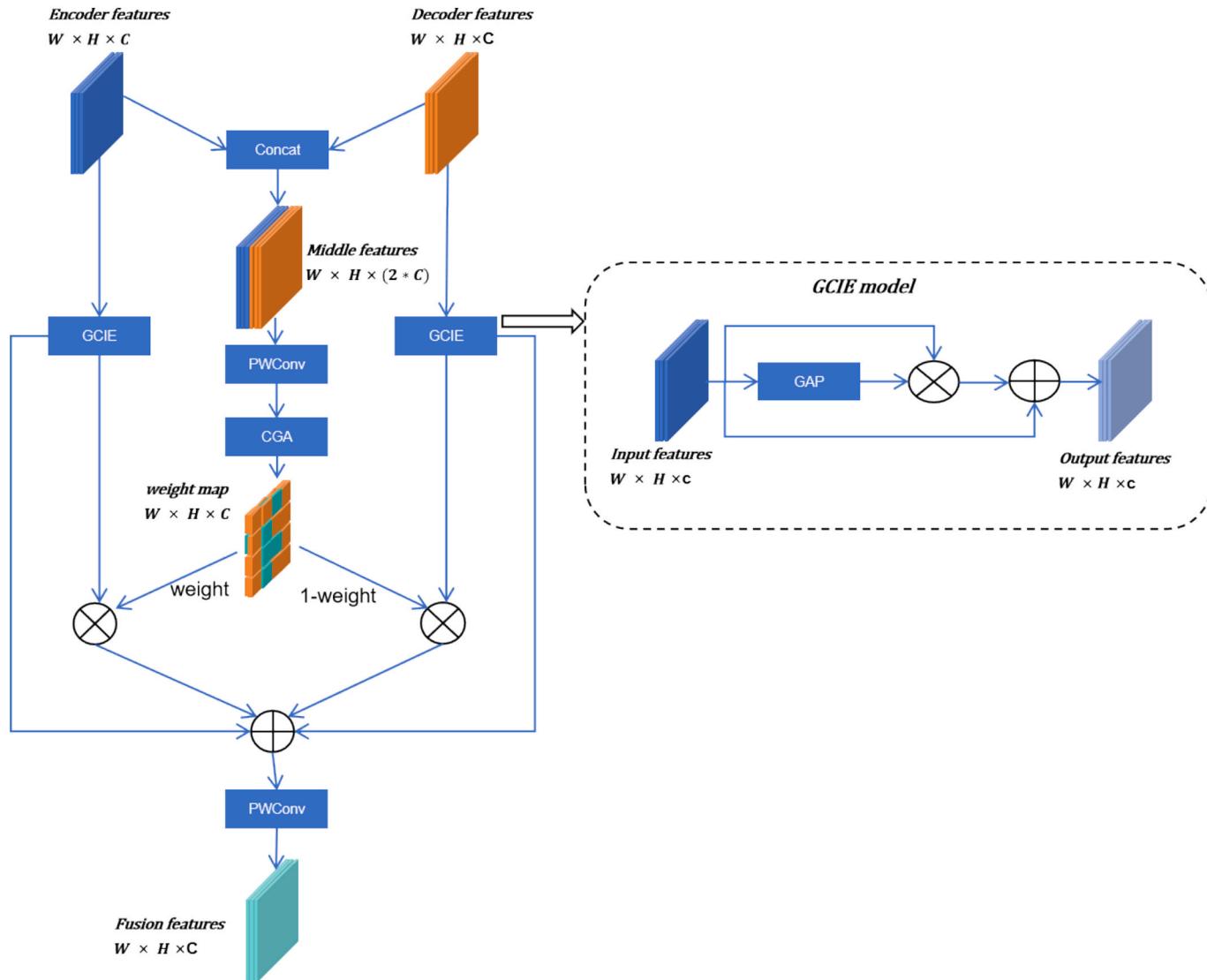


Fig. 4. Multi-scale information-guided fusion module (GCIE module).

the feature weight map Map:

$$\text{Map} = f_{CGA}(F_{middle}) \quad (13)$$

The feature weights are assigned to the input features  $F_a, F_b$  after global feature enhancement are fused and summed to obtain the final output  $F_{fusion} \in R^{C \times H \times W}$ :

$$F_{fusion} = F_a \times \text{Map} + F_b \times (1 - \text{Map}) + F_{middle} \quad (14)$$

### 3.5. Spatial group enhance and DySample

Spatial Group Enhancement (SGE) is a lightweight and effective spatial attention mechanism designed to enhance feature representations in Convolutional Neural Networks (CNNs) by introducing spatially-aware enhancement operations. The core idea of SGE is to group feature maps along the channel dimensions and augment the spatial representations of each group of feature maps with global contextual information. The mechanism is particularly effective in

capturing fine-grained spatial details while maintaining computational efficiency, thereby enabling the model to focus more accurately on the desired segmentation region and improving the performance of the segmentation network.

DySample is a rapid, effective and adaptable dynamic upsampling method. The upsampling operation is performed through the use of dynamic offset fields, and the algorithm supports two modes of operation: low-resolution processing (LP) and pixel-level processing (PL). These two modes facilitate the model's capacity to discern intricate details within the input image while preserving computational efficiency. In comparison to the conventional bilinear interpolation upsampling approach, DySample is capable of adaptively modifying the sampling operation in accordance with the characteristics of the feature map. This enables a more versatile and precise feature extraction process, ultimately facilitating the accurate delineation of the desired region.

### 3.6. Loss function

In this paper, we employ a combination of Dice loss and binary cross-entropy loss for the purpose of supervised segmentation. Binary cross-entropy loss is a versatile tool that can be utilized in a multitude of segmentation tasks. However, in scenarios where the number of foreground prostate cancer pixels is significantly smaller than that of the background pixels, the model may exhibit a tendency to favor the background, which can result in suboptimal training outcomes. The utilisation of Dice loss for the calculation of pixel similarity between two samples serves to mitigate the issue of imbalance between the number of foreground and background pixels that is prevalent in the field of medical image segmentation [24].

$$L_{\text{loss total}} = L_{\text{Dice}} + L_{\text{CE}} \quad (15)$$

## 4. Experiments and discussions

### 4.1. Experimental details

As shown in Table 1, our experiments were conducted in the Pytorch environment and performed on three 2080ti devices with 12 GB of video memory. For the learning rate configuration, we used a combination of a warm-up strategy and an exponential based learning rate decay strategy, which allows the learning rate to be dynamically adjusted throughout the model training process. The initial learning rate is set to 0.003 (due to), and stochastic gradient descent (SGD) is used as the optimizer of the model. The randomness of SGD enables it to jump out of the local optimal in the optimization process and find a better solution, which is especially critical for complex small lesion segmentation tasks. Set the momentum to 0.9 and the random seed to 42. The segmentation study of

**Table 1**  
Experimental details.

Experimental details	Prostate Cancer Segmentation	Prostate segmentation	pulmonary nodules segmentation
Experimental environment	Pytorch	Pytorch	Pytorch
Experimental equipment	3 x 2080ti	3 x 2080ti	3 x 2080ti
Initial learning rate	0.003	0.003	0.003
Optimizer	SGD	SGD	SGD
Momentum	0.9	0.9	0.9
Batchsize	4	4	16
Epoch	600	200	200
Time	8.66 h	2.5 h	2.7 h
Data set size	1086	1086	5613
Random seed	42	42	42
Average reasoning time per chart	0.245 s	0.251 s	0.065 s

prostate cancer and prostate was set to 4 batches, and the segmentation study of lung nodules was set to 16 batches. A total of 600 rounds of prostate cancer segmentation training were conducted. This value was obtained through several experiments, and if it was increased, it would enter the overfitting state. The average training time was 8.66 h. The average reasoning time per image was 0.245 s, while the training time for prostate segmentation was 2.5 h / 200 rounds, and the average reasoning time per image was 0.251 s. Over a 2.7 h training period, 200 rounds of lung nodule segmentation training were performed, with an average reasoning time of 0.065 s.

### 4.2. Prostate cancer segmentation experiment

#### a. Prostate cancer data set.

The prostate cancer dataset is a challenge for PROSTATEx. PROSTATEx [25] provides a multimodal data set of prostate cancer changes, including MRI (T2-weighted imaging, diffusion weighted imaging, etc.), along with image features and annotations extracted from these images. The dataset is derived primarily from actual prostate cancer patients and contains information related to prostate segmentation, the location of cancer lesions, and the classification of these lesions. The data in the dataset was manually screened to filter out prostate images that did not contain lesions or were too small (measuring 1–2 pixels in size and therefore not clinically significant). This process produced 1,086 images of prostate cancer changes, which were then divided into a training set and a test set in a 9:1 ratio. Before training, the data set is expanded by means of image flipping and random cropping to avoid overfitting.

#### b. Evaluation metrics.

In order to evaluate the efficacy of the prostate cancer segmentation task, four distinct metrics were employed. The evaluation metrics employed were accuracy (Acc), the Dice coefficient (Dice), the intersection over union (IOU), and the F1 score (F1 Score). The aforementioned metrics are employed in order to provide a comprehensive assessment of the performance of the segmentation model. In particular, the Accuracy metric (Acc) is comprised of a number of sub-metrics, which serve to facilitate a comprehensive evaluation of the segmentation effect across a range of regions. Firstly, Global Accuracy (G\_Acc) is employed to assess the overall classification accuracy of the model, encompassing both the foreground and background classification performance. Secondly, we also evaluated prospective accuracy (P\_Acc), which assesses the accuracy of classification of prostate cancer lesion regions, while background accuracy (B\_Acc) measures the model's segmentation effect on non-lesion regions (background). The Dice coefficient is a frequently employed metric for quantifying the degree of overlap between the segmentation results and the true annotation results. It is particularly well-suited for evaluating the accuracy of foreground region segmentation, specifically lesion region segmentation. A value approaching 1 indicates that the segmentation result is in closer alignment with the actual situation. The IOU (intersection over union) is also employed to evaluate the accuracy of segmented regions, quantifying the extent of overlap between the segmented region and the actual labelled region. The prospective IOU (P\_IOU) was calculated in order to assess the accuracy of the segmentation of regions that may potentially be prostate cancer lesions. The background IOU (B\_IOU) was calculated in order to assess the quality of the segmentation of regions that are unlikely to be prostate cancer lesions. The F1 score is a combined assessment metric based on precision and recall. It is used in the segmentation task to balance the assessment of the model's precision and recall, thus providing a further assessment of the model's overall segmentation performance. By combining these multiple evaluation metrics, an in-depth analysis of the performance of the prostate cancer segmentation model can be provided in all aspects.

### c. Experimental results.

In this study, we undertake a comprehensive comparison of the proposed MSR-Net with a number of other highly regarded segmentation methods, including Unet, FPN [26], Manet [27], Deeplabv3 [28], Deeplabv3plus [29], Unet++ [16] and Transunet [30]. Furthermore, a series of ablation experiments have been designed to validate the effectiveness of the proposed modules. For the sake of simplicity, we will refer to the fundamental U-shaped network model with a pre-trained ResNet34 backbone as the baseline method.

In the field of medical image segmentation, the Dice coefficient is the most commonly employed metric. Dice coefficients are used as a measure of ensemble similarity, with values ranging from 0 to 1. A value of 1 represents the optimal segmentation result, while a value of 0 represents the worst possible result. In this experiment, the Dice coefficient is employed as the principal evaluation metric.

As evidenced by the experimental results presented in [Table 2](#), MSR-Net demonstrates a notable enhancement in the efficacy of prostate cancer segmentation. In comparison with the optimal reference model, Unet, MSR-Net demonstrated a 4.9 % improvement in the Dice coefficient. With the exception of P\_Acc, which exhibits a slight decline in performance relative to the initial benchmark, MSR-Net demonstrates the most optimal outcomes in G\_Acc, B\_Acc, B\_IOU, P\_IOU, and F1.

As illustrated in the [Fig. 5](#), the U-Net and FPN networks exhibit significant errors in the localisation of prostate cancer boundaries and region segmentation. In contrast, the Manet, DeeplabV3, DeepLabV3plus, Unet++, and Transunet networks demonstrate notable mis-segmentation or incomplete segmentation. In comparison to the aforementioned methods, MSR-Net is capable of accurately identifying the location and boundary of prostate cancer through the utilisation of a multilayer convolutional attention mechanism (MLKA) and the fusion of multi-scale information. Concurrently, the enhanced CGA feature fusion module efficiently mitigates the impact of superfluous and inconsequential features, thereby rendering the integrated features more conducive to segmentation operations. Furthermore, the SGE module enhances the key information present in the shallow features, while the dynamic up-sampling module improves the up-sampling efficiency while maintaining the original accuracy of the features, thus achieving efficient and accurate prostate cancer segmentation [Fig. 6](#).

To further validate the effectiveness of each module, we conducted ablation experiments. The experimental results ([Table 3](#)) demonstrate a significant enhancement in the accuracy of the segmentation process through the utilisation of the ROI method, with a 10.4 % improvement in the Dice coefficient compared to the method that did not incorporate ROI. In the experiment, the MSR-Baseline with ROI method was employed, whereby both the encoder and decoder were replaced with MSR-Blocks and the network depth was set to six layers. In comparison with the Baseline, all metrics except G\_Acc demonstrate improvement, with the Dice coefficient exhibiting a 2.9 % enhancement.

In the MSR-baseline with ROI + MIGF experiment, the improved CGA feature fusion module was introduced, which resulted in further improvements to all metrics except P\_Acc. The Dice coefficient was improved by 1.1 %. Subsequently, in the MSR-baseline with ROI + MIGF

+ dynamic up-sampling experiment, the dynamic up-sampling module was added. This has the effect of significantly reducing the number of network parameters and the inference time, although the improvement in some metrics is smaller.

At last, the SGE module was incorporated into the MSR baseline through the combination of ROI, CGA, dynamic upsampling, and SGE experiments, resulting in the formation of the final MSR-Net. Following the incorporation of the SGE module, a further enhancement in segmentation accuracy was observed, accompanied by an increase of 1 % in the Dice coefficient.

In general, MSR-Net performs well in all segmentation metrics as a result of a series of optimised and fused modules, particularly in terms of the Dice coefficient.

### 4.3. Prostate segmentation experiment

#### a. Experimental data and Evaluation metrics.

The prostate dataset was also derived from the PROSTATEx challenge, which provides a dataset of multimodal prostate cancer lesions, including MRI (T2-weighted imaging, diffusion-weighted imaging, etc.). Additionally, it comprises image features and annotations extracted from these images. The dataset is primarily derived from actual prostate cancer patients and contains information pertaining to the segmentation of the prostate, the location and classification of cancer lesions. Furthermore, the metrics employ the prostate cancer segmentation metrics previously outlined (see b.Evaluation metrics in 4.1).

#### b. Experimental results.

As evidenced by the data presented in the [Table 4](#), our proposed MSR-Net methodology demonstrates notable advancements in several pivotal metrics, particularly in the Dice coefficient. In comparison with the conventional Unet approach, the Dice coefficient of MSR-Net exhibits an improvement of 2.8 percentage points (93.7 % vs. 90.9 %), thereby demonstrating enhanced segmentation precision. Furthermore, MSR-Net attained the highest values for G\_Acc (99.2 %) and B\_Acc (99.5 %) in comparison to the other models, thereby substantiating the model's superiority in terms of overall accuracy and the precision of segmentation in the background region.

Furthermore, although P\_Acc was slightly inferior to Deeplabv3 (95.9 % vs. 97.6 %), MSR-Net demonstrated superior performance in the B\_IOU (99.2 %), P\_IOU (89.3 %), and F1 (97.0 %) metrics, particularly in the P\_IOU and F1 scores. This suggests that the model not only accurately delineated the boundaries of prostate cancer, but also exhibited excellence in the prediction of the foreground region. In comparison to other competing models, methods such as Unet++ (91.6 %), Deeplabv3 (91.2 %), and FPN (91.0 %) demonstrated improvement in Dice coefficient, yet their overall performance remained inferior to that of MSR-Net, particularly in terms of key foreground intersection ratio and Dice coefficient, which are crucial metrics for segmentation tasks.

From the [Fig. 7](#), it can be observed that, due to the relatively clear boundary of the prostate, all the methods employed are able to segment

**Table 2**  
Performance comparison of Prostate Cancer Segmentation.

Methods	Dice(%)	G_Acc(%)	B_Acc(%)	P_Acc(%)	B_IOU(%)	P_IOU(%)	F1(%)
Unet [14]	55.6	95.3	96.8	<b>73.7</b>	95.1	49.6	81.9
FPN [26]	53.7	95.2	96.8	71.4	95.0	48.4	81.3
Manet [27]	51.7	95.2	97.0	68.8	95.0	47.3	80.8
Deeplabv3 [28]	52.8	95.0	96.5	72.5	94.8	47.7	80.9
Deeplabv3Plus [29]	54.1	95.0	96.4	73.2	94.7'	47.6	80.9
Unet++ [16]	50.2	95.4	97.3	66.6	95.2	47.5	80.9
Transunet [30]	43.6	91.0	92.5	68.9	90.6	32.5	72.1
Swin-Unet	51.4	90.1	95.2	71.6	93.5	46.7	79.4
Ours	<b>60.5</b>	<b>96.7</b>	<b>98.5</b>	68.8	<b>96.5</b>	<b>56.3</b>	<b>85.1</b>

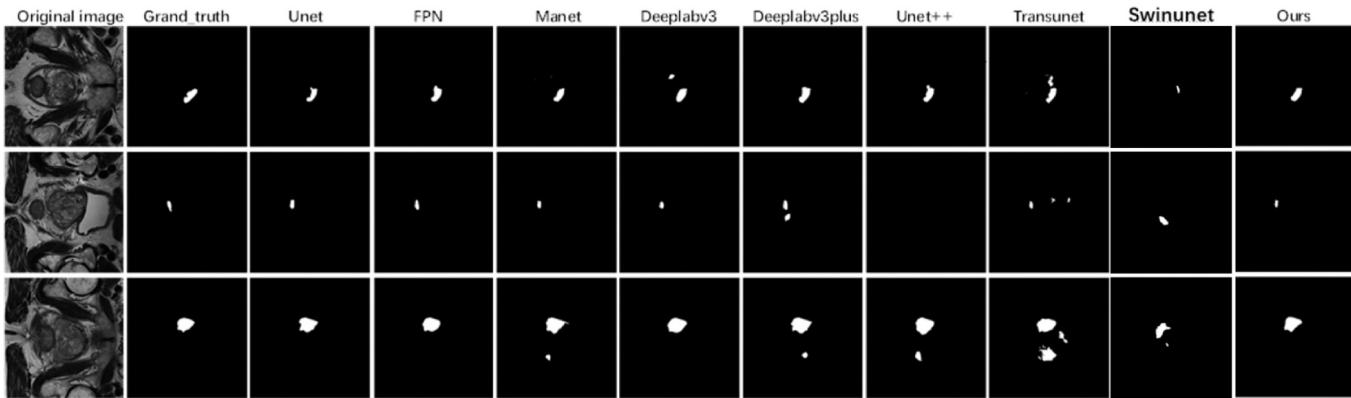


Fig. 5. The visualization results of prostate cancer segmentation.

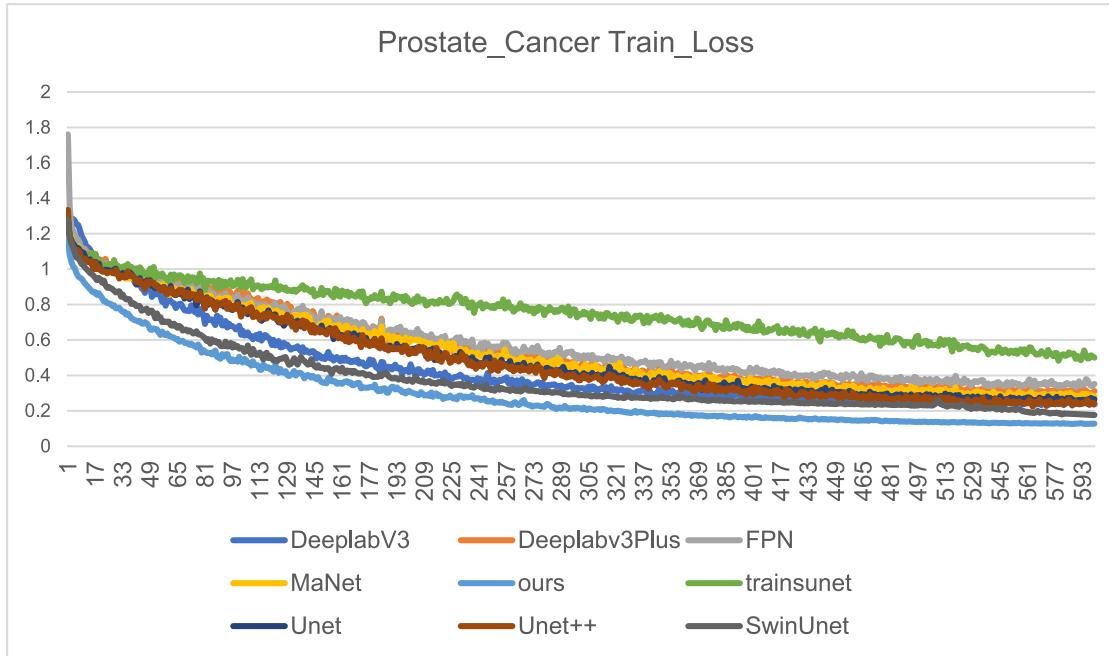


Fig. 6. Prostate cancer task segmentation Train\_loss.

**Table 3**  
Ablation Studies on Prostate Cancer Segmentation.

Methods	Dice (%)	G_Acc (%)	B_Acc (%)	P_Acc (%)	B_IOU (%)	P_IOU (%)	F1 (%)
Baseline without ROI	45.3	91.7	92.1	67.3	91.3	36.2	74.6
Baseline with ROI	55.7	95.6	97.0	72.4	95.0	50.4	82.1
MSR-baseline with ROI	58.6	95.5	97.6	73.2	95.7	53.1	83.3
MSR-baseline with ROI + MIGF	59.7	96.1	97.2	69.3	96.3	55.2	84.3
MSR-baseline with ROI + MIGF + DySample	59.5	96.6	97.8	68.4	96.7	56.1	84.7
MSR-baseline with ROI + MIGF + DySample + SGE	<b>60.5</b>	<b>96.7</b>	<b>98.5</b>	<b>68.8</b>	<b>96.5</b>	<b>56.3</b>	<b>85.1</b>

the prostate with a reasonable degree of accuracy. However, it is evident that the methods used have some degree of inconsistency in the edges, in comparison to the aforementioned methods. Our proposed MSR-Net demonstrates a superior ability to accurately segment the prostate boundary, which is indicative of its excellent performance in the task of prostate segmentation Fig. 8.

#### 4.4. Lung nodule segmentation experiment

##### a. Experimental data and Evaluation metrics.

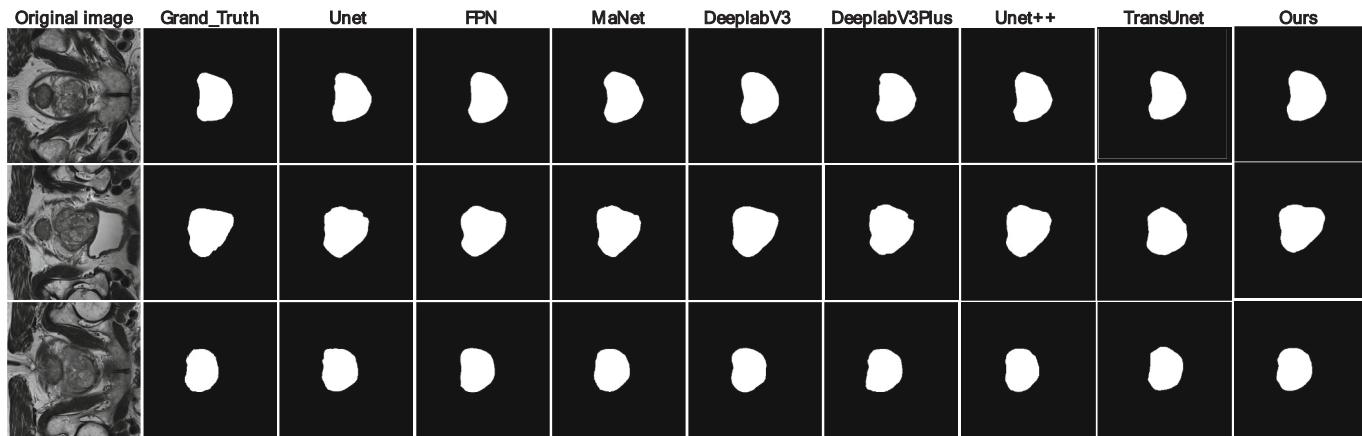
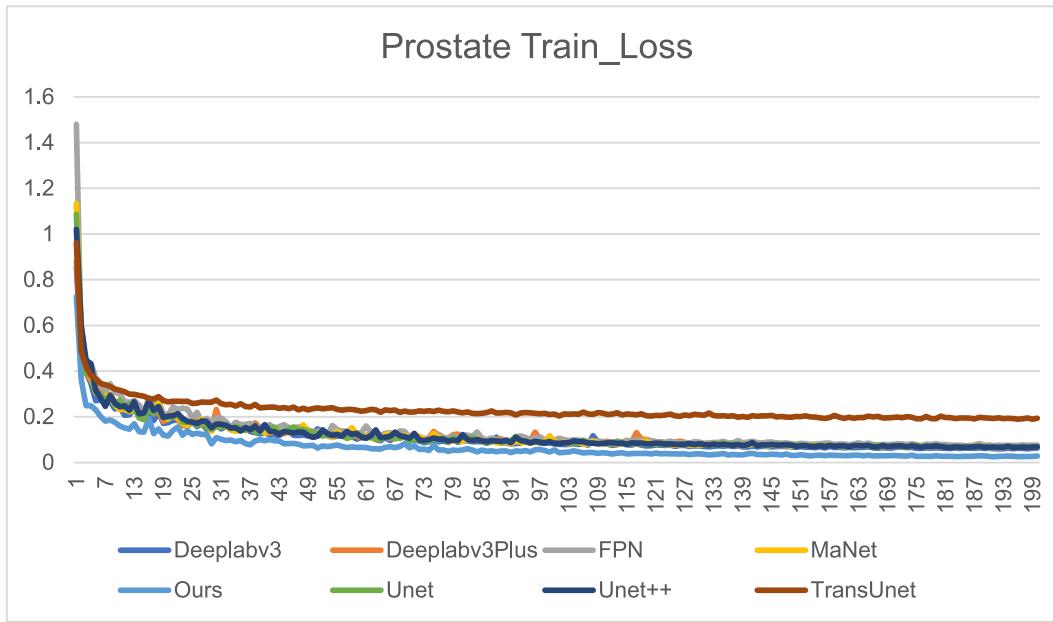
The lung nodule dataset is EDPB, which is a mixture of the public dataset LIDC and the private dataset (images of patients with lung nodules in Shanghai Electric Power Hospital). It was then manually annotated by three professional physicians and manually processed and screened. Ultimately, 5,613 images were acquired and divided into a training set and a test set for network training and testing, with an 8:2 ratio. The metrics were adopted as described above (b. Evaluation metrics in 4.1).

##### b. Experimental results.

**Table 4**

Preformance comparison of Prostate Segmentation.

Methods	Dice(%)	G_Acc(%)	B_Acc(%)	P_Acc(%)	B_IOU(%)	P_IOU(%)	F1(%)
Unet [14]	90.9	98.9	99.2	94.4	98.8	84.9	95.6
FPN [26]	91.0	99.0	99.2	96.4	98.9	86.1	96.0
Manet [27]	91.5	99.0	99.2	96.2	99.0	86.8	96.2
Deeplabv3 [28]	91.2	99.0	99.1	97.6	98.9	86.6	96.1
Deeplabv3Plus [29]	91.4	99.0	99.2	96.6	99.0	86.8	96.2
Unet++ [16]	91.6	99.0	99.2	97.0	99.0	87.1	96.3
Transunet [30]	90.1	98.7	99.1	96.4	98.7	87.3	95.8
Ours	93.7	99.2	99.5	95.9	99.2	89.3	97.0

**Fig. 7.** The visualization results of prostate segmentation.**Fig. 8.** Prostate Region Segmentation Train\_Loss.

As evidenced by the data presented in the Table 5, the MSR-Net model demonstrates notable enhancements in several pivotal performance metrics, particularly in the Dice coefficient. In comparison with the conventional Unet approach, the Dice coefficient of MSR-Net exhibits an improvement of 2.1 percentage points (76.4 % vs. 74.3 %), thereby substantiating its enhanced capacity for accurate segmentation. In comparison to other models, MSR-Net also achieved the highest values in metrics such as P\_IoU (71.3 %) and F1 (91.6 %), thereby demonstrating the model's superiority in terms of both accuracy and

overall performance with regard to the segmentation of the foreground region.

Furthermore, although MSR-Net exhibits a slight decline in P\_Acc (foreground accuracy) compared to Deeplabv3 (89.9 % vs. 90.7 %), it continues to demonstrate superior performance in core metrics such as background intersection and merger ratio (B\_IoU) and Dice coefficient, particularly in the intersection and merger ratio of foreground regions (P\_IoU) and F1 scores, which serve as a comprehensive measure of segmentation performance.

**Table 5**

Preformance comparison of Lung nodule Segmentation.

Methods	Dice(%)	G_Acc(%)	B_Acc(%)	P_Acc(%)	B_IOU(%)	P_IOU(%)	F1(%)
Unet [14]	74.3	99.7	99.8	88.7	99.7	70.1	91.1
FPN [26]	74.2	99.7	99.8	87.7	99.7	70.8	91.4
Manet [27]	74.1	99.6	99.7	86.7	99.6	66.0	89.7
Deeplabv3 [28]	65.7	99.5	99.6	90.7	99.5	61.2	87.8
Deeplabv3Plus [29]	73.3	99.7	99.8	88.8	99.7	69.3	90.8
Unet++ [16]	74.2	99.7	99.8	87.7	99.7	70.8	91.4
Transunet [30]	69.3	99.6	99.8	89.8	99.4	64.2	88.1
Ours	<b>76.4</b>	99.7	99.8	89.9	99.7	<b>71.3</b>	<b>91.6</b>

A visual inspection of the segmentation map (Fig. 9) reveals that the Unet, FPN and other networks have the tendency to segment the normal tissues into the same area, resulting in a significant discrepancy between the segmentation and the Grand\_Truth map. Furthermore, the MSR-Net segmentation is more precise, while simultaneously exhibiting enhanced accuracy in the delineation of border regions. With regard to edge details, our proposed method demonstrates a notable enhancement in comparison to other networks Fig. 10

#### 4.5. Discussions

The objective of this study is to accurately segment prostate cancer lesions. In this section, the proposed method demonstrates superior performance on the prostate cancer dataset, as evidenced by a 5.2 % improvement in Dice metrics compared to the second-best network results (60.5 % vs. 55.3 %). Furthermore, experiments were conducted on the prostate and lung nodule datasets separately. In terms of the Dice metric, there was an improvement of 2.1 % (93.7 % vs 91.6 %) on the prostate dataset and 2.2 % (76.4 % vs 74.2 %) on the lung nodule dataset. This demonstrates that our proposed network is capable of performing well on a range of tasks, indicating that it has good generalisation capabilities. The graph illustrating the segmentation results demonstrates that our network is capable of accurately determining the size and shape of lesions and tissues. Nevertheless, our method demonstrates superior performance in terms of edge detail segmentation.

However, it should be noted that our proposed segmentation network still has limitations. The present study was conducted using a single modality (T2) image, which is a relatively simple and computationally less expensive approach. Furthermore, it avoids the potential errors associated with different modal alignments. In comparison to a multimodal approach, the lack of supplementary modal information may result in reduced segmentation accuracy. Consequently, the next study will focus on the application of a multimodal approach in segmentation, with the same number of parameters or a small increase in the number of parameters, with the aim of improving the segmentation

accuracy of the network. In the case of a constant or minimal increase in the number of parameters, the objective is to enhance the accuracy of the network's segmentation. Furthermore, prostate cancer is predominantly represented as a low-frequency model on T2 images. Consequently, future research will focus on mining low-frequency information on T2 images to enhance the segmentation accuracy of prostate cancer.

#### 5. Conclusion

This paper presents a novel network for segmentation, designated MSR-Net, which is constructed based on the MSR-block, CGAfusion, SGE and DySample modules. The MSR-Net is applied for two-stage segmentation of prostate cancer. The Dice coefficient for the segmentation of prostate cancer lesions was 60.5 %. In comparison with alternative networks, the proposed network demonstrates the most optimal performance. The two-stage segmentation method proved an effective means of improving the accuracy of prostate cancer segmentation, with a 10.4 % improvement in the Dice coefficient observed in the experiment conducted with the BASELINE method. Furthermore, the network was employed in prostate and lung nodule segmentation studies. In the prostate study, the segmented data was utilized for the region of interest extraction step in the two-phase approach. Similarly, in the lung nodule study, the network demonstrated excellent performance, with a segmentation Dice reaching 93.7 % and 76.4 %, respectively. These results substantiate the generalization ability of the proposed model.

The next study will aim to enhance the precision of the network segmentation by integrating the frequency-domain attention approach, with a particular focus on the image's low-frequency information. Concurrently, a multimodal methodology is employed to capitalise on the potential of multiparametric magnetic resonance imaging, thereby facilitating more precise delineation of prostate cancer.

#### CRediT authorship contribution statement

Xinyi Chen: Writing – original draft, Visualization, Validation,

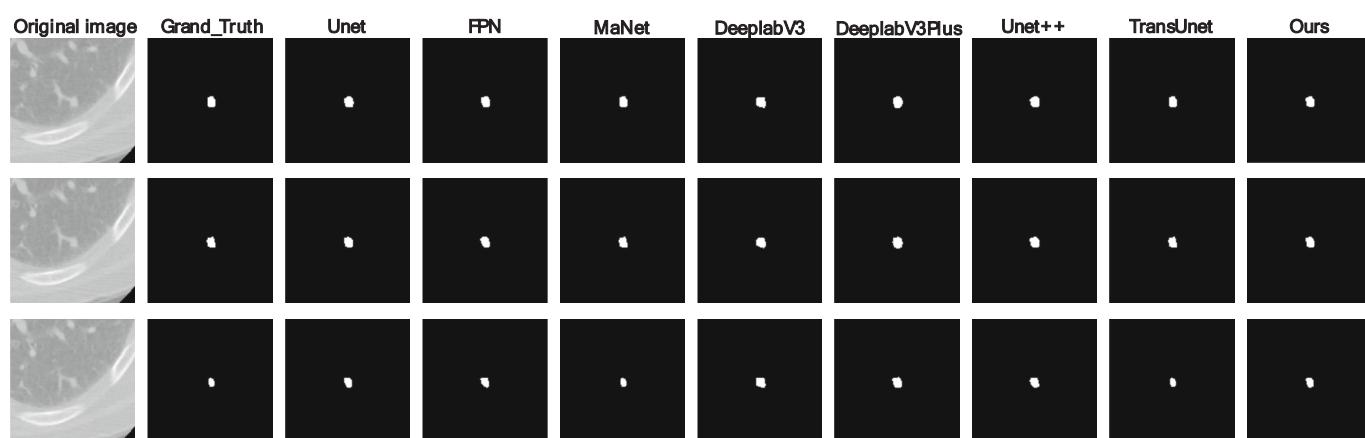
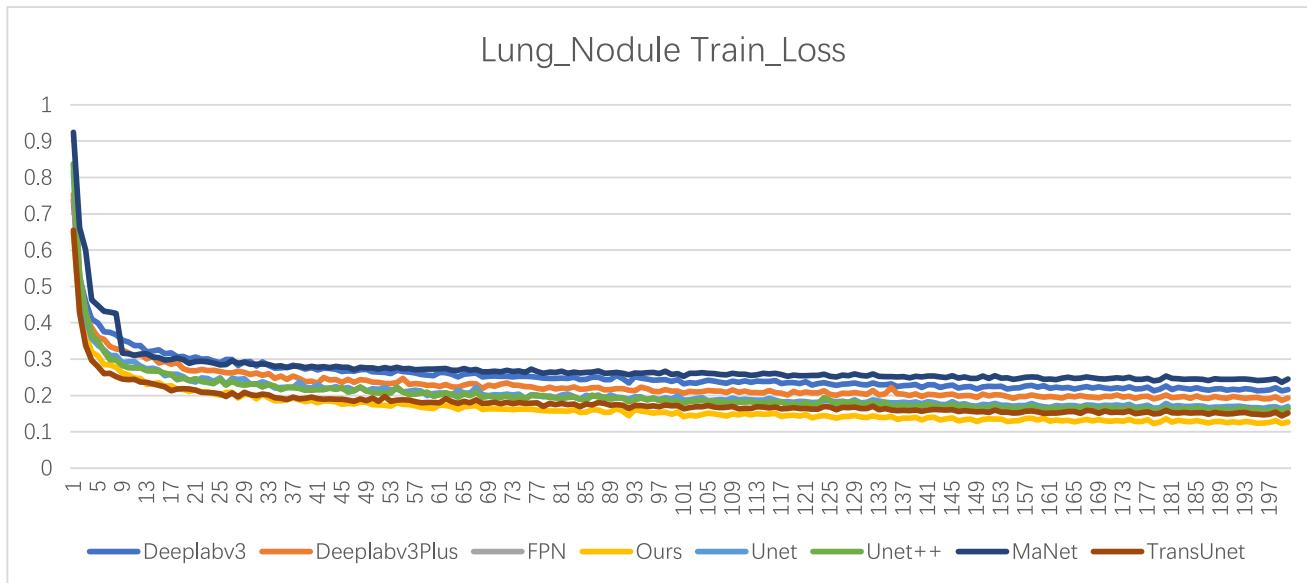


Fig. 9. The visualization results of Lung nodule segmentation.



**Fig. 10.** Lung nodule segmentation train\_loss.

Software, Project administration, Methodology, Investigation, Data curation. **Xiang Liu:** Writing – review & editing, Funding acquisition. **Yunjie Yu:** Data curation. **Yunyu Shi:** Writing – review & editing, Investigation. **Yuke Wu:** Resources, Formal analysis. **Zhenglei Wang:** Data curation. **Shuohong Wang:** Writing – review & editing, Validation.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

The data that has been used is confidential.

#### References

- [1] A. Fujihara, T. Iwata, A. Shakir, et al., Multiparametric magnetic resonance imaging facilitates reclassification during active surveillance for prostate cancer, *BJU Int.* 127 (6) (2021) 712–721.
- [2] S. Wang, X. Liu, J. Zhao, et al., Computer auxiliary diagnosis technique of detecting cholangiocarcinoma based on medical imaging: a review, *Comput. Methods Programs Biomed.* 208 (2021) 106265.
- [3] R.C. Wu, A.H. Lebastchi, B.A. Hadaschik, et al., Role of MRI for the detection of prostate cancer, *World J. Urol.* 39 (2021) 637–649.
- [4] Reda I, Ayinde B O, Elmogy M, et al. A new CNN-based system for early diagnosis of prostate cancer. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018: 207–210.
- [5] A. Comelli, N. Dahiya, A. Stefano, et al., Deep learning-based methods for prostate segmentation in magnetic resonance imaging, *Appl. Sci.* 11 (2) (2021) 782.
- [6] L. Yan, D. Liu, Q. Xiang, et al., PSP net-based automatic segmentation network model for prostate magnetic resonance imaging, *Comput. Methods Programs Biomed.* 207 (2021) 106211.
- [7] Wang Y, Li Y, Wang G, et al. Multi-scale attention network for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 5950–5960.
- [8] X. Li, X. Hu, J. Yang, Spatial group-wise enhance: Improving semantic feature learning in convolutional networks, *arXiv preprint arXiv:1905.09646* (2019).
- [9] Z. Chen, Z. He, Z.M. Lu, DEA-net: single image dehazing based on detail-enhanced convolution and content-guided attention, *IEEE Trans. Image Process.* 33 (2024) 1002–1015.
- [10] Liu W, Lu H, Fu H, et al. Learning to upsample by learning to sample. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 6027–6037.
- [11] R. Kaymak, C. Kaymak, A. Ucar, Skin lesion segmentation using fully convolutional networks: a comparative experimental study, *Expert Syst. Appl.* 161 (2020) 113742.
- [12] Janssens R, Zeng G, Zheng G. Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks. In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, 2018: 893–897.
- [13] Li X, Wang S, Wei X, et al. Fully convolutional networks for ultrasound image segmentation of thyroid nodules[C]//2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2018: 886–890.
- [14] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18. Springer international publishing, 2015: 234–241.
- [15] A. Hatamizadeh, V. Nath, Y. Tang, et al., Swin unet: swin transformers for semantic segmentation of brain tumors in mri images, in: International MICCAI brainlesion workshop, Springer International Publishing, Cham, 2021, pp. 272–284.
- [16] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]//Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4. Springer International Publishing, 2018: 3–11.
- [17] F. Isensee, J. Petersen, A. Klein, et al., nnU-Net: Self-adapting framework for u-net-based medical image segmentation (2018), *arXiv preprint arXiv:1809.10486* (1809).
- [18] Z. Tian, L. Liu, Z. Zhang, et al., PSNet: prostate segmentation on MRI based on a convolutional neural network, *J. Med. Imaging* 5 (2) (2018) 021208.
- [19] F. Song, F. Qianjian, Segmentation of prostate MR image based on deep learning network PSP-NET.Modern, *Electron Technol.* 42 (12) (2019), 148–151155.
- [20] R. Cuocolo, A. Comelli, A. Stefano, et al., Deep learning whole-gland and zonal prostate segmentation on a public MRI dataset, *J. Magn. Reson. Imaging* 54 (2) (2021) 452–459.
- [21] M. Bardis, R. Houshyar, C. Chantaduly, et al., Segmentation of the prostate transition zone and peripheral zone on MR images with deep learning, *Radiol. Imaging Cancer* 3 (3) (2021) e200024.
- [22] S. Mehralivand, D. Yang, S.A. Harmon, et al., A cascaded deep learning-based artificial intelligence algorithm for automated lesion detection and classification on biparametric prostate magnetic resonance imaging, *Acad. Radiol.* 29 (8) (2022) 1159–1168.
- [23] Y. Lin, E.C. Yilmaz, M.J. Belue, et al., Evaluation of a cascaded deep learning-based algorithm for prostate lesion detection at biparametric MRI, *Radiology* 311 (2) (2024) e230750.
- [24] S. Feng, H. Zhao, F. Shi, et al., CPFNet: context pyramid fusion network for medical image segmentation, *IEEE Trans. Med. Imaging* 39 (10) (2020) 3008–3018.
- [25] S.G. Armato III, H. Huisman, K. Drukker, et al., PROSTATEx challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images, *J. Med. Imaging* 5 (4) (2018) 044501.
- [26] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117–2125.
- [27] Liang J, Sun G, Zhang K, et al. Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 4096–4105.

- [28] L.C. Chen, G. Papandreou, F. Schroff, et al., Rethinking atrous convolution for semantic image segmentation, arXiv preprint arXiv:1706.05587 (2017).
- [29] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation, In: Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.
- [30] J. Chen, Y. Lu, Q. Yu, et al., Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306 (2021).