frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 26 June 2025
DOI 10.3389/fmed.2025.1589707

OPEN ACCESS

EDITED BY
Shaojie Tang,
Xi'an University of Posts and Telecommunications, China

REVIEWED BY
Lixia Wang,
Cedars Sinai Medical Center, United States
Chen Shanxiong,
Southwest University, China

*CORRESPONDENCE
Chengliang Wang
✉ wangcl@cqu.edu.cn
Salem Alkhalaf
✉ s.alkhalaf@qu.edu.sa

RECEIVED 07 March 2025
ACCEPTED 06 June 2025
PUBLISHED 26 June 2025

CITATION
Arshad M, Wang C, Wajeeh Us Sima M, Shaikh JA, Alkhalaf S and Alturise F (2025) RaNet: a residual attention network for accurate prostate segmentation in T2-weighted MRI. *Front. Med.* 12:1589707. doi: 10.3389/fmed.2025.1589707

RaNet: a residual attention network for accurate prostate segmentation in T2-weighted MRI

Muhammad Arshad[1], Chengliang Wang[1]*, Muhammad Wajeeh Us Sima[1], Jamshed Ali Shaikh[1], Salem Alkhalaf[2]* and Fahad Alturise[3]

[1]College of Computer Science, Chongqing University, Shapingba, Chongqing, China, [2]Department of Computer Engineering, College of Computer, Qassim University, Buraydah, Saudi Arabia, [3]Department of Cybersecurity, College of Computer, Qassim University, Buraydah, Saudi Arabia

Accurate segmentation of the prostate in T2-weighted MRI is critical for effective prostate diagnosis and treatment planning. Existing methods often struggle with the complex textures and subtle variations in the prostate. To address these challenges, we propose RaNet (Residual Attention Network), a novel framework based on ResNet50, incorporating three key modules: the DilatedContextNet (DCNet) encoder, the Multi-Scale Attention Fusion (MSAF), and the Feature Fusion Module (FFM). The encoder leverages residual connections to extract hierarchical features, capturing both fine-grained details and multi-scale patterns in the prostate. The MSAF enhances segmentation by dynamically focusing on key regions, refining feature selection and minimizing errors, while the FFM optimizes the handling of spatial hierarchies and varying object sizes, improving boundary delineation. The decoder mirrors the encoder's structure, using deconvolutional layers and skip connections to retain essential spatial details. We evaluated RaNet on a prostate MRI dataset PROMISE12 and ProstateX , achieving a DSC of 98.61 and 96.57 respectively. RaNet also demonstrated robustness to imaging artifacts and MRI protocol variability, confirming its applicability across diverse clinical scenarios. With a balance of segmentation accuracy and computational efficiency, RaNet is well suited for real-time clinical use, offering a powerful tool for precise delineation and enhanced prostate diagnostics.

KEYWORDS

RaNet, deep learning, refine feature selection, medical image segmentation, prostate cancer, feature fusion

1 Introduction

Medical image segmentation is vital for disease diagnosis, lesion localization, treatment planning, and surgical navigation. Traditional methods, requiring manual feature extraction, are often computationally expensive and lack flexibility across different clinical scenarios. In contrast, deep learning, especially convolutional neural networks (CNNs), has revolutionized segmentation by enabling end-to-end learning directly from data, eliminating the need for manual feature design (1–3). Among the deep learning architectures, U-Net (4) has gained prominence due to its efficient encoder-decoder structure, which integrates semantic features for precise segmentation. Recent

advancements, including dilated convolutions (5), attention mechanisms (6), and multi-scale feature extraction (7), have further enhanced U-Net's capabilities, addressing challenges like ambiguous boundaries and complex anatomical structures. Despite these advancements, CNNs still struggle with modeling long-range spatial dependencies, which are critical for resolving complex structures, especially in medical images with subtle contrast variations and adjacent tissue influences. This limitation has prompted the exploration of transformer-based models (8, 9), which excel at capturing global context, improving segmentation accuracy. However, incorporating long-range dependencies remains an ongoing challenge for segmentation tasks where contextual information is essential.

Prostate cancer is one of the leading causes of cancer-related deaths in men, making accurate prostate segmentation on T2-weighted magnetic resonance imaging crucial for diagnosis, treatment planning, and disease monitoring. While manual segmentation by radiologists is the gold standard, it is time-consuming and prone to inter-observer variability. Automated prostate segmentation has proven difficult due to complex anatomical textures, subtle contrast differences, and imaging artifacts such as motion blur and inhomogeneity (10, 11). Despite advancements in MRI technology, automated prostate segmentation remains challenging due to the region's complex anatomical textures, subtle contrast variations with adjacent tissues, and imaging artifacts such as motion blur or inhomogeneity. While machine learning (ML) and deep learning (DL) methods have shown promise, conventional models often fail to capture the PZ's intricate patterns (12). Standard convolutional neural networks (CNNs), constrained by fixed-size kernels, struggle to model long-range spatial dependencies essential for resolving ambiguous boundaries (13). Furthermore, variability in MRI acquisition protocols, patient anatomy, and artifact profiles limits the model's robustness to MRI artifacts and generalizability across different clinical settings, as demonstrated by the experimental results in later sections (14). Despite significant advancements in medical image segmentation, several challenges remain. Existing segmentation models struggle to capture complex patterns and subtle variations in medical images, particularly in challenging regions like the prostate. Additionally, they often fail to effectively focus on relevant regions and refine feature maps at multiple scales, leading to reduced accuracy. Variability in semantic and spatial information from multiple decoder layers also limits segmentation performance.

To address these limitations, we propose RaNet, a novel ResNet50-based framework enhanced with dynamic attention mechanisms. By integrating attention modules into skip connections, RaNet adaptively prioritizes relevant regions in T2-weighted MRI, suppressing noise while refining feature extraction for precise prostate localization. This approach aims to overcome the shortcomings of fixed-receptive-field CNNs and variability-induced performance degradation, offering a robust solution for accurate and generalizable prostate segmentation.

The key contributions of this paper are as follows:

- Introduce a robust encoder-decoder architecture that utilizes residual connections to effectively extract hierarchical features while preserving spatial details, allowing the model to capture complex patterns and subtle variations in medical images, particularly in the prostate.
- Propose a novel MSAF that acts as an attention mechanism at skip connections, enabling the model to focus on relevant regions and refine feature maps at multiple scales. This refinement improves segmentation accuracy and reduces errors.
- Integrate a FFM that fuses outputs from multiple decoder layers using bilinear upsampling. This module effectively combines both semantic and spatial information, leading to more accurate and robust segmentation results.
- Extensive experiments on a large MRI dataset demonstrate the superior performance of the proposed model, achieving a Dice similarity coefficient of 0.92. The model also reduces false positive and false negative rates, outperforming conventional segmentation methods.

This article is structured into several key sections that address our research on prostate segmentation. Section 2 reviews related work, highlighting various modular approaches. Section 3 outlines our automated segmentation framework, detailing its innovative components. Section 4 describes the dataset, implementation details, and evaluation metrics, Section 5 along with experimental results supported by ablation studies and explain the comparative studies with other state-of-the-art methodologies. In Section 6, critically discuss the findings and their implications within the existing literature. Finally, Section 7 summarizes the key insights and suggests directions for future research.

2 Related work

Prostate segmentation, particularly in T2-weighted MRI, presents significant challenges due to complex anatomical textures, subtle contrast variations, and the influence of surrounding tissues. In recent years, various advancements in deep learning architectures have been proposed to address these challenges and improve segmentation accuracy. The related work in this area is categorized into three key sections: encoder architectures, attention mechanisms in skip connections, and multi-scale feature fusion. These approaches have contributed significantly to enhancing the robustness and performance of prostate segmentation models, as outlined in the sections below.

2.1 Encoder architectures for medical imaging

The success of deep learning in medical segmentation is anchored in encoder-decoder architectures such as U-Net (4), which employs a symmetric structure to hierarchically extract features through its contracting path (encoder) and reconstruct precise segmentations via its expanding path (decoder), bridged by skip connections to retain spatial details. Building on this foundation, He et al. (15) introduced ResNet50, which addresses vanishing gradients in deep networks through residual blocks

that enable stable training by learning residual mappings via shortcut connections. Recent adaptations of ResNet variants for prostate MRI segmentation have further optimized this backbone: Gurkan et al. (16) integrated ResNet50 as the encoder within a Mask R-CNN framework, leveraging its feature reuse capabilities to improve segmentation of anatomically distinct prostate zones by aligning region proposals with high-resolution feature maps. Similarly, Li et al. (17) enhanced ResNet50's context capture by replacing standard convolutions with dilated convolutions in deeper layers, strategically expanding the network's receptive fields without increasing computational overhead a critical advantage for resolving subtle intensity variations in T2-weighted MRI. Talaat et al. (18) integrates ResNet50 with Faster R-CNN and dual optimizers, aiming to improve prostate cancer detection accuracy. This model demonstrates significant performance improvements in detecting and localizing prostate lesions in MRI images. Another study focuses on utilizing ResNet50 for feature extraction, achieving high classification accuracy in detecting prostate cancer lesions, and providing a robust solution for automated cancer detection (19). Furthermore, the MM-UNet architecture combines a modified ResNet50 encoder with Mamba blocks, refining feature extraction and enhancing segmentation precision in prostate MRI scans. This model not only improves segmentation accuracy but also aids in resolving complex boundaries within the prostate zone (20). These targeted modifications demonstrate how ResNet50's residual learning principles can be tailored to balance model depth and efficiency while preserving the precision required for prostate zonal anatomy segmentation.

2.2 Attention mechanisms in skip connections

Attention gates refine the propagation of features between the encoder and decoder by adaptively suppressing irrelevant regions in skip connections. The seminal Attention U-Net (21) introduced channel-wise attention mechanisms, where feature maps from the encoder are weighted using attention coefficients derived from the decoder's higher-level features, enabling the model to focus on salient regions like pancreatic tumors while ignoring background noise. For prostate MRI segmentation, Nash et al. (22) extended the use of attention mechanisms by incorporating spatial attention into skip connections. This approach generates pixel-wise attention maps that focus on anatomically plausible boundaries of the prostate, leveraging learned spatial correlations between the encoder and decoder features. Building on this, Zhang et al. (23) further improved robustness against MRI intensity inhomogeneity by integrating self-attention mechanisms with deformable convolutions in the skip connections. The self-attention captures long-range dependencies to resolve ambiguous edges, while deformable convolutions adjust the receptive fields to better accommodate irregular prostate shapes. Federico et al. (24) proposes Long-Range 3D Self-Attention to capture multi-scale features in MRI scans, improving segmentation accuracy. Another introduces a pseudo-3D Global–Local Channel Spatial Attention mechanism to enhance segmentation of prostate zones in T2-weighted MRI, significantly improving accuracy for both the transition and peripheral zones (25). Building on these advancements, our hybrid channel-spatial attention modules unify channel-wise and spatial attention within skip connections, dynamically amplifying discriminative prostate features (e.g., subtle intensity gradients) while suppressing confounding background signals through joint optimization of channel relevance and spatial saliency.

2.3 Multi-scale feature fusion

Effective fusion of hierarchical features is critical for segmenting small, ambiguous structures like the prostate, where local texture details and global anatomical context must be cohesively integrated. Zhao et al. (26) introduced Feature Pyramid Networks (FPNs), which merge multi-scale encoder outputs through lateral connections, creating a pyramid of features that combines high-resolution shallow layers (rich in spatial details) with semantically strong deeper layers (capturing contextual information). For prostate MRI segmentation, Santhirasekaram et al. (27) adapted this approach by incorporating geometric constraints into the fusion process, penalizing segmentations that violate anatomical priors (e.g., irregular prostate topology) through a loss term that enforces smoothness and connectivity in the fused feature maps. Li et al. (28) further advanced multi-scale fusion by introducing learnable weights to dynamically adjust the contribution of ResNet50's intermediate features during aggregation, enabling the model to prioritize scales resilient to common MRI artifacts such as motion blur or intensity inhomogeneity emphasize multi-scale feature fusion for prostate segmentation in MRI. AGMSF-Net integrates a multi-scale attention mechanism and 3D transformer module, improving segmentation accuracy and achieving a DSC of 93.68% on a local dataset (29). Another approach uses a multistream fusion encoder with spatial attention maps, enhancing accuracy, particularly for small lesions, and achieving improved performance on the ProstateX dataset (30).

3 Methodology

The RaNet model (Figure 1), designed for medical image segmentation tasks, follows an encoder-decoder architecture enhanced with attention mechanisms and feature fusion techniques. The encoder consists of convolutional blocks with residual connections to preserve spatial integrity while extracting high-level features. The attention mechanism, implemented through the MSAF, refines feature maps by focusing on relevant regions using varying kernel sizes and dilation rates, improving segmentation accuracy. The decoder mirrors the encoder with deconvolutional layers and skip connections to retain essential spatial details for accurate boundary delineation. Additionally, the Feature Fusion Module (FFM) merges outputs from multiple decoder layers using bilinear upsampling, ensuring robust and precise segmentation. The final segmentation mask is computed by combining these refined features, supporting accurate diagnosis and treatment planning. This architecture optimizes feature retention and relevance, significantly enhancing segmentation

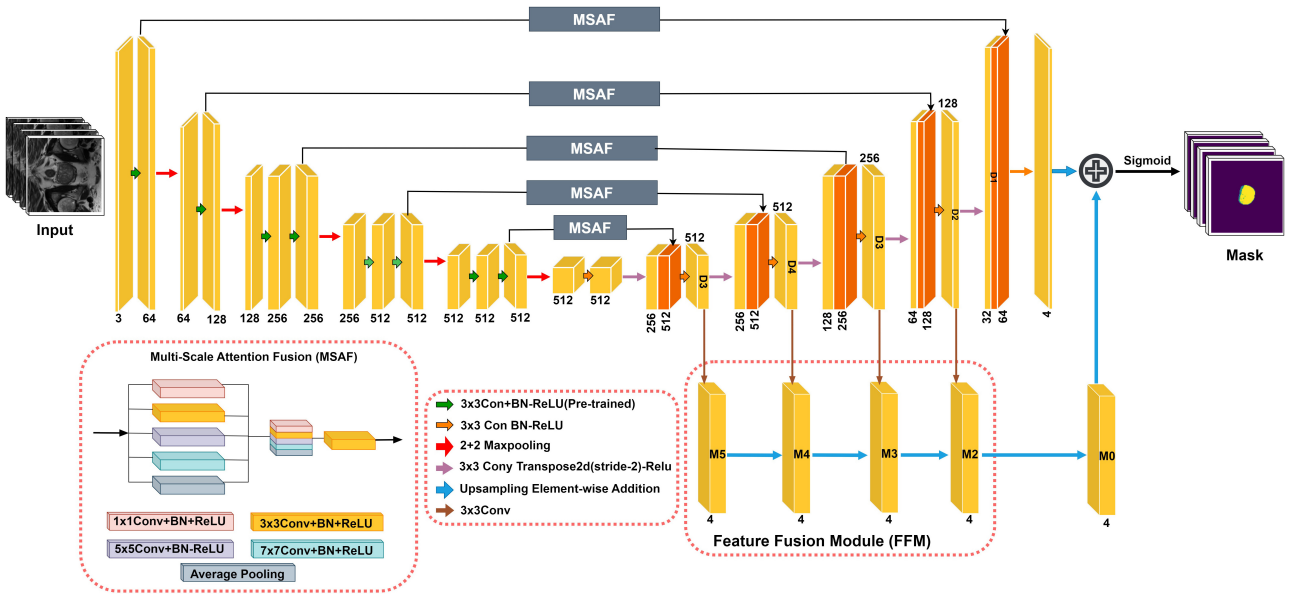

FIGURE 1
Proposed RaNet framework with MSAF and FFM on left and right respectively.

TABLE 1 Summary of model components.

Component	Description
DCNet	Modified ResNet50 with dilated convolutions and removed initial max-pooling.
MSAF	Multi-scale attention fusion with varying kernel sizes and dilation rates.
FFM	Bilinear upsampling to merge decoder outputs, enhancing segmentation accuracy.
Fusion strategy	Combines outputs from multiple layers to retain both spatial and semantic info.
Loss function	Hybrid loss combining Dice, weighted cross-entropy, and boundary loss.

performance in medical imaging. Overall component summary of the RaNet is given in the Table 1.

3.1 DilatedContextNet

In this paper, we enhance the ResNet50 architecture by making three key modifications to the encoder (the feature extraction part) of the network. These modifications aim to preserve spatial information, improve feature extraction for segmentation tasks, and introduce uncertainty modeling. In the original ResNet50, a max-pooling layer is applied immediately after the first convolutional layer. However, max-pooling reduces the spatial resolution of the feature maps, which is undesirable for segmentation tasks where fine-grained spatial details are critical. To address this issue, we remove the initial max-pooling layer, preserving the spatial dimensions of the input image as it passes through the network. Let the input to the encoder be represented as:

$$X_0 \in \mathbb{R}^{H_0 \times W_0 \times C_0} \tag{1}$$

where H_0, W_0, and C_0 represent the height, width, and number of channels of the input image. In the original ResNet50, the max-pooling operation would reduce the spatial dimensions as:

$$X_1 = \text{MaxPool}(X_0) \tag{2}$$

However, in the modified version, we remove this step and directly pass X_0 through the first convolutional layer:

$$X_1 = \text{Conv}(X_0) \tag{3}$$

Thus, we avoid any early downsampling, allowing the network to retain the original spatial resolution.

The bottleneck block in the fourth layer has a stride of 2, which reduces the spatial resolution of the feature maps too aggressively for segmentation. To address this, we replace the stride-2 bottleneck block with a regular convolutional block with stride 1. This ensures that the spatial resolution is maintained in the fourth layer, allowing the network to retain more detailed features.

The regular convolution operation is mathematically represented as:

$$X_2 = \text{Conv}(X_1, \text{stride} = 1) \tag{4}$$

where X_1 is the output from the previous layer and the stride is set to 1 to preserve spatial dimensions.

To further improve the ability of the network to capture larger contextual information while maintaining spatial resolution, we introduce a dilated convolution in the second block of the fourth layer. Dilated convolutions increase the receptive field by introducing gaps between the convolutional kernel's elements,

allowing the network to capture larger contextual information without downsampling the feature maps.

The dilated convolution operation with dilation rate r is expressed as:

$$\text{DilatedConv}(X, r) = \sum_{i=1}^{k} w_i X_{i \cdot r} \tag{5}$$

where w_i represents the convolutional kernel, $X_{i \cdot r}$ represents the dilated input, and r is the dilation rate. By using a dilated convolution in the second block of the fourth layer, we prevent spatial information loss while capturing a broader context.

Thus, the output from the dilated bottleneck block is:

$$X_3 = \text{DilatedConv}(X_2, r) \tag{6}$$

where X_2 is the output from the previous regular convolutional block.

To transform the DCNet into a Bayesian neural network, we incorporate dropout layers after each block. Dropout is a regularization technique that randomly deactivates a proportion of neurons during training, helping the model generalize better by preventing overfitting. In a Bayesian context, dropout can be interpreted as a way to approximate the posterior distribution of the network's weights, making the network more robust and capable of modeling uncertainty.

Mathematically, the output of a neuron with dropout is given by:

$$\hat{y}_i = \begin{cases} y_i & \text{with probability } p \\ 0 & \text{with probability } (1 - p) \end{cases} \tag{7}$$

where p is the probability of keeping the neuron active (typically $p = 0.5$).

For each block in the encoder, dropout is applied as follows:

$$X_i' = \text{Dropout}(X_i, p) \tag{8}$$

where X_i represents the output from the i-th block, and p is the dropout rate.

This DCNet architecture aim to enhance prostate segmentation, crucial for prostate cancer diagnosis. Removing the initial max-pooling layer preserves fine spatial details essential for accurate boundary detection. Replacing the stride-2 bottleneck with a regular convolution and introducing dilated convolutions retain spatial resolution while expanding the receptive field to capture broader tissue context. Incorporating dropout layers converts the model into a Bayesian neural network, improving generalization and robustness.

3.2 Multi-scale attention fusion

The architecture of the MSAF is depicted in Figure 1. During the feature map learning phase from the encoder, convolution operations are performed simultaneously before being directly connected to the decoder. The resulting feature maps are then concatenated. To reduce the number of channels, a transition block consisting of a convolution layer followed by a ReLU activation function (Conv + PreLU) is added. Various methods for concatenating convolutional layers with different kernel sizes and dilation rates were explored. The optimal MSAF architecture, chosen from experimental comparisons and shown in Figure 1, incorporates a 1 × 1 Conv + PreLU block, a 3 × 3 Conv + PreLU block with dilation rate 1, a 3 × 3 Conv + PreLU block with dilation rate 2, a 3 × 3 Conv + PreLU block with dilation rate 3, and an image pooling layer. The overall procedure of MSAF can be mathematically expressed as:

$$\begin{aligned} X_N = f(\text{Concat}(K_1(X^O), K_3(X^O, \text{D} = 1), K_3(X^O, \text{D} = 2), \\ K_3(X^O, \text{D} = 3), G_I(X^O))) \end{aligned} \tag{9}$$

where X_O represents the features from the encoder, X_N denotes the new features generated by MSAF, which are then passed to the decoder. K_i refers to the convolution operations with kernel size i, and GI stands for the global image pooling operation. The function $f(\cdot)$ is the transition operation that adjusts the number of channels in X_N to match that of X_O.

The MSAF structure bears similarities to the X-ception module, which provides several advantages, such as the ability to capture multi-scale information through different convolution kernel sizes. This is particularly useful for real-world applications where target segmentation requires handling scale variance. Additionally, as demonstrated in GoogleNet, the use of multiple kernel sizes can facilitate faster convergence by decomposing sparse matrices into dense matrix operations. To further enhance the model's performance, a global image average pooling operation is incorporated, which has been shown to effectively capture global context information in several studies.

3.3 Feature fusion module

In U-Net-based models, probability maps are generated by the final layer of the decoder. Since feature maps in convolutional networks cannot simultaneously retain both semantic and spatial information, robust and accurate probability maps are obtained by fusing the outputs from different decoder layers. This process can be formulated as follows:

$$F_o = \sum_{i=2}^{5} U_u(F_i) \tag{10}$$

where F_o represents the fused feature map generated by the FFM, and F_i corresponds to the feature map produced by decoder layer D_i (for $i = 2, 3, 4, 5$). The function U_u denotes bilinear upsampling, which is used to adjust the probability map size to match the original image dimensions.

The final segmentation mask is computed as:

$$\text{Mask} = \text{Sigmoid}(F_1 + F_o) \tag{11}$$

where F_1 is the feature map from the last decoder layer D_1, and the segmentation mask is obtained by applying the sigmoid function to the sum of F_1 and F_o.

3.4 Loss function

To address the challenges of segmenting the prostate in MRI, which exhibits ambiguous boundaries and class imbalance between foreground and background, we design a hybrid loss function. The total loss $\mathcal{L}_{\text{total}}$ combines three components: a region-based Dice loss ($\mathcal{L}_{\text{Dice}}$), a distribution-aware weighted cross-entropy loss ($\mathcal{L}_{\text{WCE}}$), and a boundary-focused loss ($\mathcal{L}_{\text{boundary}}$) to refine edge delineation. The combined loss is defined as:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{Dice}} + \lambda_2 \mathcal{L}_{\text{WCE}} + \lambda_3 \mathcal{L}_{\text{boundary}}, \tag{12}$$

where λ_1, λ_2, and λ_3 are weighting coefficients balancing the contributions of each term.

Dice loss ($\mathcal{L}_{\text{Dice}}$) The Dice loss (31) mitigates class imbalance by maximizing the overlap between the predicted segmentation mask $\hat{y}$ and ground truth y:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_i y_i \hat{y}_i + \epsilon}{\sum_i y_i + \sum_i \hat{y}_i + \epsilon} \tag{13}$$

where ϵ is a smoothing factor to avoid division by zero.

Weighted cross-entropy loss ($\mathcal{L}_{\text{WCE}}$) To penalize misclassifications in underrepresented prostate regions, we use a weighted cross-entropy loss (4):

$$\mathcal{L}_{\text{WCE}} = -\sum_i w \cdot \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right] \tag{14}$$

where w is a class weight inversely proportional to the foreground pixel frequency.

Boundary loss ($\mathcal{L}_{\text{boundary}}$) To improve segmentation accuracy along poorly defined prostate edges, we adopt a boundary loss (32) that computes the symmetric Euclidean distance between the contours of y and $\hat{y}$:

$$\mathcal{L}_{\text{boundary}} = \sum_{p \in \partial y} \min_{q \in \partial \hat{y}} \|p - q\|_2 + \sum_{q \in \partial \hat{y}} \min_{p \in \partial y} \|q - p\|_2 \tag{15}$$

where ∂y and $\partial \hat{y}$ denote the contours of the ground truth and prediction, respectively.

4 Dataset and pre-processing

4.1 Dataset

4.1.1 PROMISE12

The prostate MRI data used in this study are sourced from the PROMISE12 Challenge dataset (33). The PROMISE12 Challenge dataset is designed for studying MRI prostate segmentation and comprises 50 T2 MRI scans of the prostate region from 50 patients. The data were collected from multiple hospitals, ensuring representation of a clinical setting with diverse vendors and acquisition protocols. Specifically, the dataset includes 50 MRI volumes with corresponding training labels and 30 MRI volumes for testing, which lack ground truth images. A few image samples are shown in Figure 2.

4.1.2 ProstateX

The ProstateX dataset, originally part of the PI-CAI dataset, did not include segmentation masks for anatomical regions. However, a subsequent study by Cuocolo et al. (34) annotated 204 cases from the original ProstateX dataset, providing both lesion masks and anatomical region masks. This enhanced dataset is valuable for research in lesion detection and anatomical region segmentation in prostate imaging. Example images from this dataset are shown in Figure 2.

4.2 Evaluation metrics

The performance of the proposed RaNet model for segmenting the prostate zones in T2-weighted MRI is evaluated using the Dice coefficient, Intersection over Union (IoU), and accuracy. The Dice coefficient measures the similarity between the predicted segmentation mask of the prostate zones and the ground truth mask, while IoU assesses the overlap between these masks. Accuracy represents the proportion of correctly predicted pixels in the segmentation task.

$$\text{DSC} = \frac{2 \cdot \text{intersection}(A, B)}{\text{size}(A) + \text{size}(B)} \tag{16}$$

$$\text{IoU} = \frac{\text{intersection}(A, B)}{\text{union}(A, B)} \tag{17}$$

$$\text{Accuracy} = \frac{\text{intersection}(A, B)}{\text{union}(A, B)} \tag{18}$$

The Dice Similarity Coefficient (DSC) quantifies the overlap between the predicted prostate zone segmentation mask and the ground truth mask A and B, considering both false positives and false negatives. The Intersection Over Union (IoU) measures the proportion of overlap between the predicted and ground-truth masks, normalized by their union. Accuracy measures the fraction of correctly identified pixels in the segmentation task by comparing the intersection of the sets with their union.

4.3 Implementation details

The experiments were conducted on a single NVIDIA RTX 3090 GPU using the PyTorch framework. The input images were resized to 256 × 256 pixels, and pixel intensities and voxel resolutions were standardized to enhance model generalizability. The data was split into 70% for training, 20% for validation, and 10% for testing. Both datasets underwent comprehensive preprocessing, which included a structured hierarchy of techniques, such as various augmentation strategies (no augmentation, vertical, horizontal, and diagonal shifts) and a mix of original and downsampled data. To optimize model training, hyperparameters were fine-tuned by exploring batch sizes (4, 8, and 16) and learning rates (1×10^{-2}, 1×10^{-3}, and 1×10^{-4}). Each combination was evaluated using metrics like accuracy, precision, specificity, DSC and Jaccard to assess its effectiveness. The augmentation

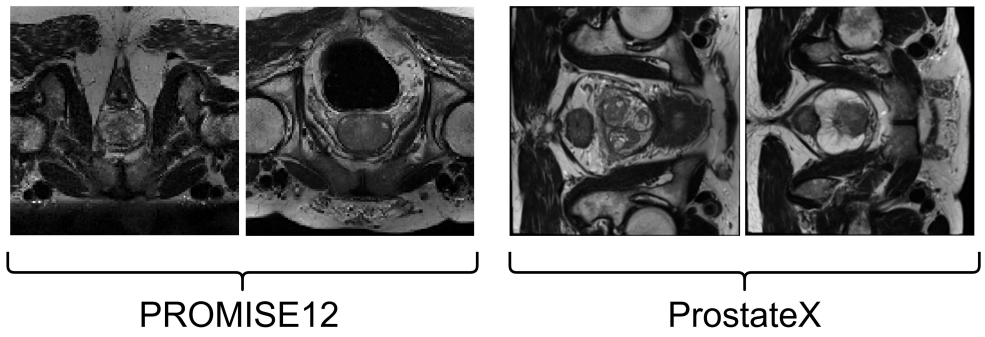

FIGURE 2
Image samples from PROMISE12 dataset **(on Left)**, ProstateX dataset **(on Right)**.

TABLE 2 Augmentation and preprocessing parameters for the experiment.

Attributes	Description
Horizontal shift	True
Zoom	True
Vertical shift	True
Diagonal shift	True
Hue saturation	True
Random brightness	True
Random contrast	True

and preprocessing parameters are given in Table 2. Additionally, preprocessing metrics for the PROMISE12 dataset are provided in Table 3, while those for the ProstateX dataset are outlined in Table 4. The model was trained end-to-end using the Adam optimizer (35). A ReduceLROnPlateau learning rate scheduler dynamically adjusted the learning rate during training. The model was trained for a maximum of 200 epochs with early stopping based on validation loss. To address the challenges of prostate segmentation, a hybrid loss function was employed to manage class imbalance and improve edge delineation.

5 Results and analysis

5.1 Comparison with different encoders

The comparison of different encoders with our proposed modules on the PROMISE12 dataset shows significant changes in performance across various models (Table 5). Starting with ResNet-18, the baseline model, we observe an improvement with ResNet-34, which shows a DSC increase of 1.67%, IoU improvement of 1.65%, and an accuracy boost of 2.5%. These improvements are attributed to ResNet-34's deeper architecture, which allows for better feature extraction through residual connections. Further improvements are seen with ResNet-50, which achieves a DSC increase of 0.88%, IoU improvement of 0.24%, and a slight decrease in accuracy of 0.25%. This increase in DSC and IoU can be attributed to the enhanced feature extraction capabilities and deeper layers of ResNet-50, allowing it to capture more detailed spatial features. The EfficientNet models show more varied results. EfficientNetB3 has a slight decrease in DSC (down by 0.75%) and a small increase in specificity (up by 0.43%) compared to ResNet-50, but its accuracy and IoU are lower than ResNet-50. This suggests that while EfficientNetB3 is more efficient, it may not be able to capture the same level of detail as deeper models like ResNet-50. EfficientNetB4, on the other hand, achieves a DSC increase of 0.44%, IoU improvement of 0.36%, and accuracy increase of 0.57%, indicating that a deeper and more optimized EfficientNet performs well while maintaining computational efficiency. EfficientNetB5 shows a small improvement in DSC (up by 0.20%) and IoU (up by 0.12%), while its accuracy increases by 0.54% compared to ResNet-50. Finally, RaNet outperforms all models, with a DSC increase of 1.69% over EfficientNetB5, an IoU improvement of 0.43%, and a significant accuracy gain of 1.30%. This improvement is due to RaNet's use of dilated convolutions, attention mechanisms, and the removal of max-pooling layers, which enhance feature retention, preserve spatial resolution, and refine segmentation performance.

On the ProstateX dataset (Table 6), RaNet also demonstrates notable improvements and variations in performance. Starting with ResNet-18, the baseline model, we observe a DSC improvement of 2.75%, IoU increase of 2.59%, and accuracy gain of 1.75% with ResNet-34. This improvement is due to ResNet-34's deeper architecture, which allows for better feature extraction and handling of more complex patterns through residual connections. Further increases are observed with ResNet-50, which shows a DSC increase of 0.68%, IoU improvement of 0.74%, and accuracy improvement of 0.09% compared to ResNet-34. The additional layers in ResNet-50 refine feature extraction and improve segmentation, capturing more detailed and complex spatial information. In contrast, the EfficientNet models show varying performance. EfficientNetB3 underperforms compared to ResNet-50, with a DSC decrease of 2.09% and IoU decrease of 4.05%. This decrement is likely due to EfficientNetB3's compact design, which prioritizes computational efficiency but may not capture as detailed features as deeper models like ResNet-50. EfficientNetB4, however, shows a slight improvement over ResNet-50, with a DSC increase of 0.68%, IoU increase of 0.74%, and accuracy increase of 0.54%. This indicates that a deeper and

TABLE 3 Performance evaluation with different pre-processing and hyper-parameter settings on PROMISE12.

			Accuracy	Precision	Specificity	DSC	Jaccard
Pre-processing	Data augment	None	95.81	94.61	94.23	92.57	87.92
		Vertical	96.54	95.12	95.58	93.39	88.71
		Horizontal	96.89	95.49	96.14	95.55	89.19
		Diagonal	98.21	95.88	95.95	97.66	89.96
Hyper-parameters	Batch	4	97.83	93.79	94.15	92.21	86.65
		8	98.19	94.42	97.91	92.73	87.37
		16	98.78	95.44	95.71	93.77	88.25
	Learning rate	1×10^{-2}	97.63	91.32	95.75	95.82	86.47
		1×10^{-3}	99.61	92.76	97.91	97.54	87.83
		1×10^{-4}	98.51	92.19	94.23	94.62	86.84

TABLE 4 Performance evaluation with different pre-processing and hyper-parameter settings on ProstateX.

			Accuracy	Precision	Specificity	DSC	Jaccard
Pre-processing	Data augment	None	94.82	90.38	89.13	90.28	83.29
		Vertical	96.43	91.61	92.72	91.94	85.97
		Horizontal	95.37	91.17	92.20	91.32	84.23
		Diagonal	97.71	92.52	93.89	92.65	86.48
Hyper-parameters	Batch	4	95.46	89.91	89.37	90.13	85.28
		8	92.73	91.57	91.53	91.42	86.65
		16	96.51	91.67	91.19	92.76	87.88
	Learning Rate	1×10^{-2}	93.85	91.24	91.49	90.71	84.72
		1×10^{-3}	97.44	90.62	93.37	93.17	86.58
		1×10^{-4}	94.14	92.29	92.86	91.34	86.73

TABLE 5 Comparison with different encoders on PROMISE12.

Model	DSC%	Sensitivity%	Specificity%	Accuracy%	IoU%
ResNet-18	94.37	86.71	93.46	96.07	95.47
ResNet-34	96.04	88.89	95.38	98.57	97.12
ResNet-50	96.92	87.16	96.10	98.32	97.36
EfficientNetB3	96.17	86.94	96.29	97.15	95.84
EfficientNetB4	95.61	87.25	97.54	97.89	96.72
EfficientNetB5	96.32	88.39	96.67	98.43	97.26
RaNet	98.61	89.42	98.90	99.73	97.69

more optimized EfficientNet model benefits from better feature extraction while maintaining efficiency. EfficientNetB5 shows a small decrement in DSC (down by 1.52%) and IoU (down by 0.79%), likely due to diminishing returns as model depth increases without corresponding performance gains.

Finally, RaNet outperforms all models, achieving a DSC increase of 0.95%, IoU improvement of 0.74%, and an accuracy gain of 0.86%. This can be attributed to RaNet's architecture, which utilizes dilated convolutions, removes max-pooling layers, and integrates attention mechanisms, resulting in superior feature retention and refined segmentation performance.

5.2 Ablation study

The ablation study demonstrates the performance improvement with each module added to the base UNet architecture (Table 7). Starting with the base UNet model, it

TABLE 6 Comparison with different encoders on ProstateX.

Model	DSC%	Sensitivity%	Specificity%	Accuracy%	IoU%
ResNet-18	92.19	84.63	92.17	96.56	92.30
ResNet-34	94.94	86.72	94.80	98.31	94.89
ResNet-50	95.62	87.43	96.35	98.40	95.42
EfficientNetB3	93.53	84.59	93.16	94.38	91.37
EfficientNetB4	95.29	86.41	96.28	96.94	94.56
EfficientNetB5	94.10	86.68	95.83	96.61	94.63
RaNet	96.57	87.49	96.74	97.26	95.16

provides solid performance, achieving a DSC of 94.82% and accuracy of 95.52% on PROMISE12, and a DSC of 93.16% and accuracy of 93.17% on ProstateX. The first improvement comes with the addition of DCNet, which modifies the ResNet50 backbone by removing the initial max-pooling layer and replacing the stride-2 bottleneck with a regular convolution. These modifications preserve spatial resolution and enhance the model's ability to capture finer details. The inclusion of DCNet results in an increase in DSC of 1.49% and an improvement in precision of 0.91% on PROMISE12, while on ProstateX, the DSC increases by 1.70% and the accuracy by 2.49%. Adding the MSAF module introduces multi-scale feature fusion, allowing the model to capture features at varying scales. This module improves the model's ability to focus on relevant regions with greater precision, leading to improved segmentation of complex structures like the prostate. MSAF results in an increase in DSC of 1.35% and an improvement in accuracy of 2.46% in PROMISE12, and in ProstateX, the DSC increases by 0.93% and the accuracy by 1.07%. The final module, FFM merges the outputs from multiple decoder layers using bilinear upsampling, ensuring that both spatial and semantic information are preserved. This module strengthens the segmentation by consolidating the features learned at different levels of the network. The addition of FFM leads to a DSC increase of 0.95% and an accuracy boost of 0.84% on PROMISE12, and on ProstateX, DSC increases by 0.78% and accuracy by 0.53%. In conclusion, the integration of DCNet, MSAF, and FFM results in significant performance improvements, with RaNet achieving a DSC of 98.61% and accuracy of 99.73% on PROMISE12, and a DSC of 96.57% and accuracy of 97.26% on ProstateX, demonstrating the effectiveness of the combined architecture for prostate zone segmentation.

TABLE 7 Ablation performance comparison between PROMISE12 and ProstateX.

Models	PROMISE12		ProstateX	
	DSC%	Accuracy%	DSC%	Accuracy%
UNet	94.82	95.52	93.16	93.17
UNet + DCNet	96.31	96.43	94.86	95.66
UNet + DCNet + MSAF	97.66	98.89	95.79	96.73
UNet + DCNet + MSAF + FFM	98.61	99.73	96.57	97.26

5.3 Comparison with state-of-the-art methods

The performance comparison of RaNet with state-of-the-art prostate segmentation models on the PROMISE12 dataset (Table 8) reveals significant improvements. Compared to the original UNet (4), RaNet achieves a DSC increase of 3.79%, an IoU improvement of 4.93%, and an accuracy boost of 4.21%. These gains are attributed to RaNet's use of dilated convolutions, attention mechanisms, and the removal of max-pooling layers, which help preserve spatial resolution and refine feature maps. When compared to MicroSeg-Net (10), which uses multi-scale feature fusion and attention mechanisms, RaNet shows a DSC increase of 3.09%, IoU improvement of 3.12%, and accuracy increase of 2.60%, highlighting its superior feature retention and spatial resolution. Against PZS-Net (36), which incorporates a pyramid structure and attention layers, RaNet demonstrates a DSC increase of 2.32% and IoU improvement of 1.64%, thanks to its more effective attention-based fusion and feature extraction capabilities. Finally, RaNet outperforms nnUNet (37), which is known for its adaptive design and strong performance, by showing a DSC improvement of 1.66%, an IoU increase of 1.21%, and an accuracy boost of 1.13%. This performance is due to RaNet's ability to capture both fine details and larger contextual information through its dilated convolutions and attention mechanisms. Overall, RaNet outperforms all SOTA models, with significant improvements in DSC, IoU, and accuracy, demonstrating the effectiveness of its architecture in prostate zone segmentation. Comparison with other previous work for PROMISE12 dataset is given in Table 9.

Table 10 presents the performance comparison of RaNet with state-of-the-art prostate segmentation models on the ProstateX dataset, showing clear improvements. Compared to the original UNet (4), RaNet achieves a DSC increase of 3.41%, an IoU improvement of 4.90%, and an accuracy boost of 4.09%. These significant improvements stem from RaNet's architectural innovations, such as dilated convolutions, attention mechanisms, and the removal of max-pooling layers, which preserve spatial resolution and refine feature maps. When

TABLE 8 State-of-the-art comparison on PROMISE12.

Model	DSC%	Sensitivity%	Specificity%	Accuracy%	IoU%
UNet (4)	94.82	84.23	91.11	95.52	92.76
MicroSeg-Net (10)	95.52	87.83	94.27	97.13	94.57
PZS-Net (36)	96.29	88.91	96.89	98.99	96.05
nnUNet (37)	96.95	87.97	97.11	98.60	96.48
RaNet	98.61	89.42	98.90	99.73	97.69

TABLE 9 Comparison with previous methods on Promise12.

References	Technique	Dataset	Cases	DSC%
Jia et al. (38)	3D APA-Net	PROMISE12, ASPS13	140	90.10
Zhu et al. (39)	BOWDA-Net	PROMISE12, BWH	146	92.54
Wang et al. (40)	SegDGAN	PROMISE12, Decathlon, ISBI13, QIN-PROSTATE	335	91.66
Qian et al. (41)	ProSegNet	PROMISE12, ProstateX	80	90.80
Meyer et al. (42)	Multi-Stream-CNN	PROMISE12, In-house dataset, ProstateX	19	93.00
Ocal et al. (43)	Triple Fusion Model	PROMISE12, NCI-ISBI 2013	80	91.90
Jia et al. (44)	MSD-Net	PROMISE12, 12CVB, NCI-ISBI13	180	92.90
Chen et al. (45)	RASEU-Net	PROMISE12, Private Dataset	30	80.70
Li et al. (28)	DRCU-Net	PROMISE12	80	91.60
Bhandary et al. (46)	nnU-Net	PROMISE12, Medical Segmentation Decathlon	50	91.20
Ma et al. (47)	ResGNet	PROMISE12, Prostate158, NCI-ISBI13, PI-CAI	1,764	94.40
Ours	RaNet	PROMISE12	50	98.61

TABLE 10 State-of-the-art comparison on ProstateX.

Model	DSC%	Sensitivity%	Specificity%	Accuracy%	IoU%
UNet (4)	93.16	82.19	90.14	93.17	90.26
MicroSeg-Net (10)	93.92	85.42	94.27	94.56	91.86
PZS-Net (36)	95.16	85.96	95.54	95.91	93.43
nnUNet (37)	95.82	86.64	96.18	96.37	94.07
RaNet	96.57	87.49	96.74	97.26	95.16

compared to MicroSeg-Net (10), which uses multi-scale feature fusion and attention mechanisms, RaNet shows a DSC increase of 2.65%, IoU improvement of 3.30%, and accuracy increase of 2.70%, highlighting RaNet's superior spatial resolution and attention-based fusion strategy. In comparison to PZS-Net (36), which incorporates pyramid structures and attention layers, RaNet demonstrates a DSC increase of 1.41% and an IoU improvement of 1.73%, benefiting from its more refined attention-based fusion and feature retention capabilities. Finally, RaNet outperforms nnUNet (37), achieving a DSC improvement of 0.75%, IoU increase of 1.09%, and accuracy boost of 0.89%. The performance gains of RaNet are attributed to its advanced architecture, which captures both fine details and larger contextual information through dilated convolutions and attention mechanisms. Overall, RaNet demonstrates superior performance across all metrics, with substantial improvements over all other SOTA models, highlighting the effectiveness of its architecture for prostate zone segmentation on the ProstateX dataset. Comparison with other previous work for ProstateX dataset is given in Table 11.

Figures 3, 4 compares the segmentation results from UNet, MicroSeg-Net, PZS-Net, nnUNet, and RaNet. While UNet serves as a strong baseline, it struggles with incomplete regions and imprecise boundaries due to its lack of spatial attention and feature refinement. MicroSeg-Net improves feature map refinement but still fails to capture complete regions due to inadequate spatial resolution preservation. PZS-Net offers better region visualization but suffers from blurry boundaries due to insufficient refinement in the decoding phase. nnUNet, though comparable to RaNet in segmentation performance, faces challenges with accurately delineating zonal boundaries. In contrast, RaNet outperforms all models, delivering sharper, more complete segmentations with well-defined boundaries. This is attributed to RaNet's architectural innovations, such as the removal of max-pooling in ResNet50, dilated convolutions to preserve spatial integrity, MSAF for

TABLE 11 Comparison with previous methods on ProstateX.

References	Technique	Dataset	Cases	DSC%
Yu et al. (48)	SPCT	ProstateX	914	92.23
Zhong et al. (49)	ProSegDiff	ProstateX, NCI-ISBI, PROMISE12	-	89.16
Qian et al. (41)	ProSegNet	ProstateX, PROMISE12	346	89.20
Liu et al. (50)	CriDiff	ProstateX, NCI-ISBI	-	87.40
Yan et al. (51)	CCT-Unet	ProstateX, Huashan dataset	200	80.30
Hung et al. (52)	CAT-nnU-Net	ProstateX, private dataset	193	83.90
Nguyen and Fernandez-Quilez (53)	nnU-Net	ProstateX	204	98.00
Wei et al. (54)	attention U-Net	ProstateX, Prostate158, MSD	443	82.00
Nai et al. (55)	HighRes3DNet	ProstateX	160	89.00
Ours	RaNet	ProstateX	204	96.57

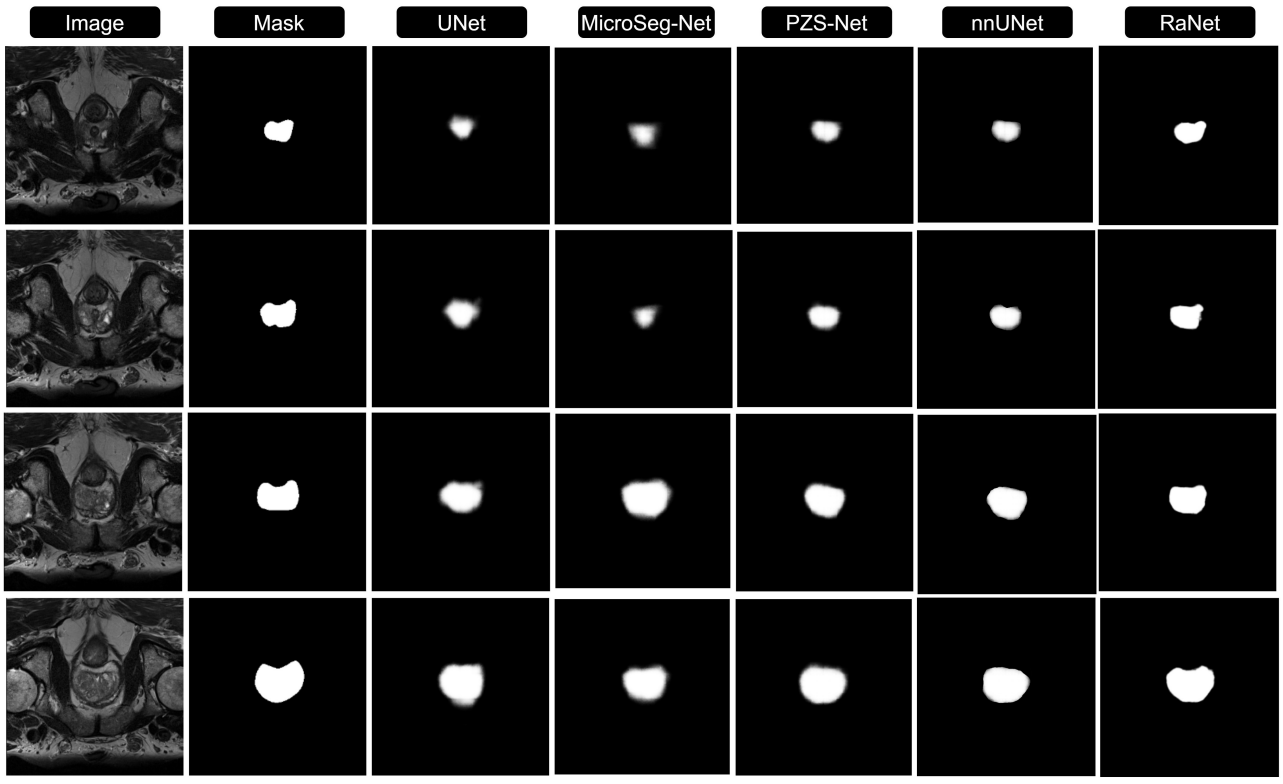

FIGURE 3
Comparison of RaNet's segmentation performance with different state-of-the-art methods on PROMISE12 dataset. From left to right: input image, ground truth, UNet, MicroSeg-Net, PZS-Net, nnUNet, and RaNet.

focusing on relevant regions, and FFM for refining the final segmentation. These enhancements make RaNet particularly effective in complex medical image segmentation tasks, ensuring high accuracy and precise boundary delineation.

6 Discussion

This study introduces RaNet, a novel model for prostate segmentation in MRI images, which outperforms traditional and state-of-the-art segmentation models, including UNet, MicroSeg-Net, PZS-Net, and nnUNet. The architecture of RaNet integrates several innovations, such as the DCNet encoder, MSAF, and FFM, which collectively enhance segmentation performance. RaNet achieved a DSC of 98.61% on the PROMISE12 dataset and 96.57% on the ProstateX dataset, demonstrating superior accuracy compared to the existing models. The DCNet encoder plays a crucial role in preserving spatial resolution by eliminating the initial max-pooling layer and using dilated convolutions to expand the receptive field. These modifications help RaNet

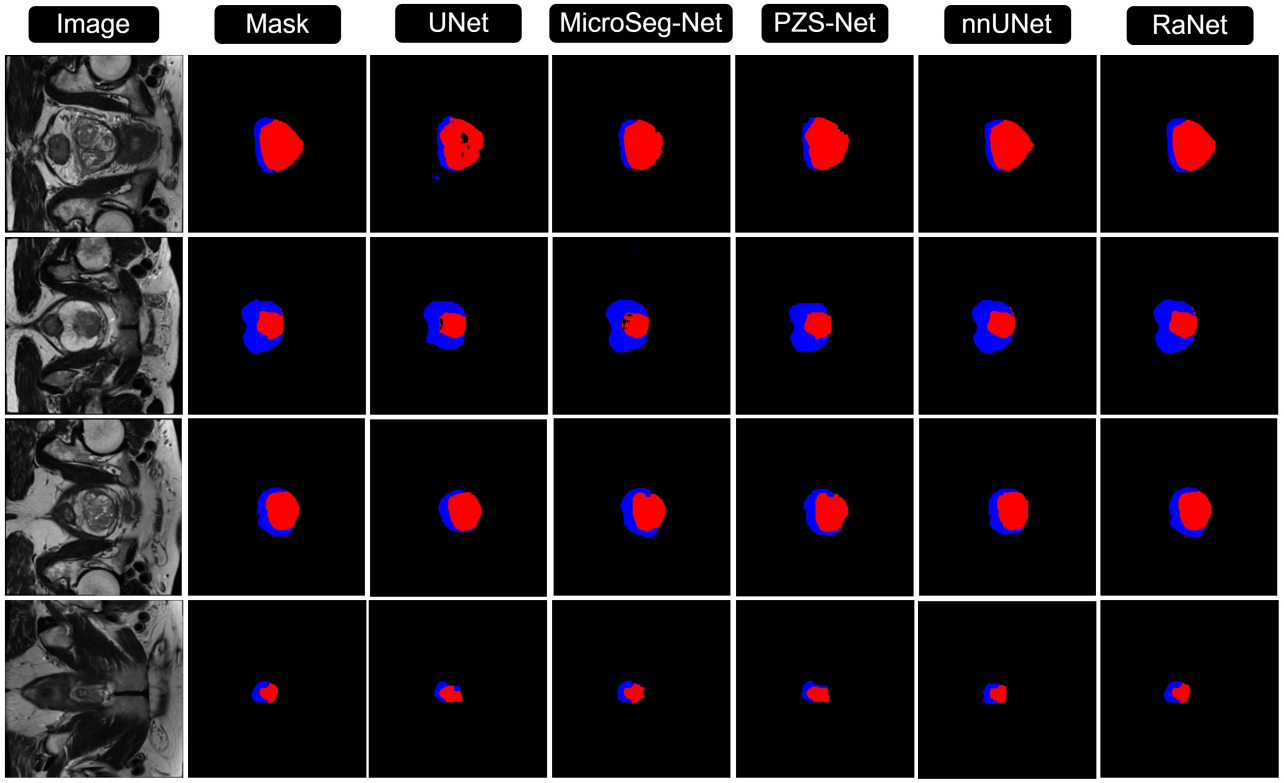

FIGURE 4
Comparison of RaNet's segmentation performance with different state-of-the-art methods on ProstateX dataset. From left to right: input image, ground truth, UNet, MicroSeg-Net, PZS-Net, nnUNet, and RaNet.

capture both fine details and larger contextual features, crucial for accurate prostate segmentation. MSAF, with its multi-scale feature fusion mechanism, refines the model's ability to focus on relevant regions at multiple scales, further improving segmentation accuracy. Finally, the FFM consolidates features from various decoder layers using bilinear upsampling, ensuring robust and precise segmentation output by maintaining both spatial and semantic information. RaNet's performance was compared with other state-of-the-art models, including nnUNet, which is known for its adaptive design and strong performance across various datasets. While nnUNet performed well, RaNet outperformed it on the ProstateX dataset, achieving a DSC improvement of 0.75% and an IoU improvement of 1.09%. This superior performance can be attributed to RaNet's architectural enhancements, such as dilated convolutions and attention mechanisms, which enable more precise feature retention and better boundary delineation, particularly for the complex prostate boundaries.

7 Conclusion

RaNet represents a significant advancement in prostate segmentation, surpassing existing models in both accuracy and efficiency. The integration of a deep ResNet-based encoder, attention mechanisms, and optimized pooling strategies enables RaNet to achieve superior performance, even with smaller datasets. These results suggest RaNet could be a valuable tool in clinical applications for prostate cancer diagnosis, improving diagnostic accuracy and treatment planning. Accurate segmentation of prostate regions is essential for clinicians to develop effective treatment strategies, as precise delineation directly influences outcomes. RaNet's high accuracy, even when trained on limited data, makes it a promising solution for real-world medical scenarios with scarce annotated data. Despite its advantages, RaNet faces challenges in high-resolution image processing and requires further optimization for computational efficiency. Additionally, extensive validation in clinical settings is needed to ensure its generalizability across diverse patient populations and imaging protocols. While RaNet performs well with smaller datasets, its computational demand could be a limitation for widespread use in real-time clinical scenarios. Future work will focus on improving RaNet's computational efficiency, particularly for high-resolution medical images, ensuring that inference speed is optimized while maintaining accuracy for real-time clinical use. Additionally, expanding its application to other medical imaging tasks, such as tumor detection and organ segmentation, will enhance its clinical relevance and solidify its role in medical image analysis and clinical decision-making.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://promise12.grand-

challenge.org/; https://www.cancerimagingarchive.net/collection/prostatex/.

Author contributions

MA: Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing. CW: Conceptualization, Project administration, Supervision, Validation, Writing – review & editing. MW: Formal analysis, Investigation, Writing – review & editing. JS: Software, Visualization, Writing – review & editing. SA: Funding acquisition, Resources, Supervision, Writing – review & editing. FA: Formal analysis, Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2025).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. Generative AI is used for the writing proof reading.

Publisher's note

References

1. Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: a survey. *IET Image Process*. (2022) 16:1243–67. doi: 10.1049/ipr2.12419

2. Mienye ID, Swart TG, Obaido G, Jordan M, Ilono P. Deep convolutional neural networks in medical image analysis: a review. *Information*. (2025) 16:195. doi: 10.3390/info16030195

3. Oubaalla A, El Moubtahij H, El Akkad N. Medical image segmentation using deep learning: a survey. In: Motahhir S, Bossoufi B, editors. *Digital Technologies and Applications*. Cham: Springer Nature Switzerland (2023). p. 974–83. doi: 10.1007/978-3-031-29860-8_97

4. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. (2015). doi: 10.1007/978-3-319-24574-4_28

5. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv:1511.07122*. (2016).

6. Xie Y, Yang B, Guan Q, Zhang J, Wu Q, Xia Y. Attention mechanisms in medical image segmentation: a survey. *arXiv:2305.17937*. (2023).

7. Pan P, Zhang C, Sun J, Guo L. Multi-scale conv-attention U-Net for medical image segmentation. *Sci Rep*. (2025) 15:12041. doi: 10.1038/s41598-025-96101-8

8. Pu Q, Xi Z, Yin S, Zhao Z, Zhao L. Advantages of transformer and its application for medical image segmentation: a survey. *Biomed Eng Online*. (2024) 23:14. doi: 10.1186/s12938-024-01212-4

9. Feng Y, Cong Y, Xing S, Wang H, Ren Z, Zhang X. GCFormer: multi-scale feature plays a crucial role in medical images segmentation. *Knowl-Based Syst*. (2024) 300:112170. doi: 10.1016/j.knosys.2024.112170

10. Jiang H, Imran M, Muralidharan P, Patel A, Pensa J, Liang M, et al. MicroSegNet: a deep learning approach for prostate segmentation on micro-ultrasound images. *Comput Med Imag Graph*. (2024) 112:102326. doi: 10.1016/j.compmedimag.2024.102326

11. Fassia MK, Balasubramanian A, Woo S, Vargas HA, Hricak H, Konukoglu E, et al. Deep learning prostate MRI segmentation accuracy and robustness: a systematic review. *Radiology*. (2024) 6:e230138. doi: 10.1148/ryai.230138

12. Zaev RI, Romanov AY, Solovyev RA. Segmentation of prostate cancer on TRUS images using ML. In: *2023 International Russian Smart Industry Conference (SmartIndustryCon)*. IEEE (2023). doi: 10.1109/SmartIndustryCon57312.2023.10110727

13. Zhang Y, Zhou C, Guo S, Wang C, Yang J, Yang Z, et al. Deep learning algorithm-based multimodal MRI radiomics and pathomics data improve prediction of bone metastases in primary prostate cancer. *J Cancer Res Clin Oncol*. (2024) 150:78. doi: 10.1007/s00432-023-05574-5

14. Pan X, Wang S, Liu Y, Wen L, Lu M. iPCa-Former: a multi-task transformer framework for perceiving incidental prostate cancer. *IEEE Signal Process Lett*. (2024) 31:785–9. doi: 10.1109/LSP.2024.3372787

15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2016). doi: 10.1109/CVPR.2016.90

16. Gürkan C, Budak A, Karatas H, Akín K. Segmentation of prostate zones on a novel MRI database using mask R-CNN: an implementation on PACS system. *J Faculty Eng Architect*. (2024) 11:507. doi: 10.17341/gazimmfd.1153507

17. Han L, Xiao S, Li Z, Li H, Zhao X, Han Y, et al. Enhanced self-supervised learning for multi-modality MRI segmentation and classification: a novel approach avoiding model collapse. *arXiv Preprint arXiv.2407.10377*. (2024).

18. Talaat FM, El-Sappagh S, Alnowaiser K, Hassan E. Improved prostate cancer diagnosis using a modified ResNet50-based deep learning architecture. *BMC Med Inform Decis Mak*. (2024) 24:23. doi: 10.1186/s12911-024-02419-0

19. Pinkham DW, Sala IM, Soisson ET, Wang B, Deeley MA. Are you ready for a cyberattack? *J Appl Clin Med Phys*. (2021) 22:4–7. doi: 10.1002/acm2.13422

20. Du Q, Wang L, Chen H, A. mixed Mamba U-net for prostate segmentation in MR images. *Sci Rep*. (2024) 14:19976. doi: 10.1038/s41598-024-71045-7

21. Oktay O, Schlemper J, Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: learning where to look for the pancreas. *arXiv:1804.03999*. (2018).

22. Zaridis DI, Mylona E, Tachos N, Kalantzopoulos CN, Marias K, Tsiknakis M,et al. ResQu-Net: effective prostate's peripheral zone segmentation leveraging the representational power of attention-based mechanisms. *Biomed Signal Process Control*. (2024) 93:106187. doi: 10.1016/j.bspc.2024.106187

23. Wang S, Wang Y, Peng Y, Chen X. MSA-Net: multi-scale feature fusion network with enhanced attention module for 3D medical image segmentation. *Comput Electric Eng*. (2024) 120:109654. doi: 10.1016/j.compeleceng.2024.109654

24. Pollastri F, Cipriano M, Bolelli F, Grana C. Long-range 3D self-attention for MRI prostate segmentation. In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. (2022). p. 1–5. doi: 10.1109/ISBI52829.2022.9761448

25. Krishnan C, Onuoha E, Hung A, Sung KH, Kim H. Multi-attention mechanism for enhanced pseudo-3D prostate zonal segmentation. *J Imag Inform Med*. (2025) 2025, 1–12. doi: 10.1007/s10278-025-01401-0

26. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. *arXiv:1612.01105*. (2017). p. 2881–2890. doi: 10.1109/CVPR.2017.660

27. Santhirasekaram A, Winkler M, Rockall A, Glocker B, A. geometric approach to robust medical image segmentation. *Med Image Anal*. (2024) 97:103260. doi: 10.1016/j.media.2024.103260

28. Liu Y, Zhu Y, Wang W, Zheng B, Qin X, Wang P. Multi-scale information residual network: Deep residual network of prostate cancer segmentation based on multi scale information guidance. *Biomed Signal Process Control*. (2025) 110:108132. doi: 10.1016/j.bspc.2025.108132

29. Li Y, Wu Y, Huang M, Zhang Y, Bai Z. Attention guided multi scale feature fusion network for automatic prostate segmentation. *Comput Mater Continua*. (2024) 78:1649–68. doi: 10.32604/cmc.2023.046883

30. Jiang M, Yuan B, Kou W, Yan W, Marshall H, Yang Q, et al. Prostate cancer segmentation from MRI by a multistream fusion encoder. *Med Phys*. (2023) 50:5489–504. doi: 10.1002/mp.16374

31. Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE (2016). p. 565–571. doi: 10.1109/3DV.2016.79

32. Kervadec H, Bouchtiba J, Desrosiers C, Granger E, Dolz J, Ayed IB. Boundary loss for highly unbalanced segmentation. *Med Image Anal*. (2021) 67:101851. doi: 10.1016/j.media.2020.101851

33. Litjens G, Toth R, van de Ven W, Hoeks C, Kerkstra S, van Ginneken B, et al. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *Med Image Anal*. (2014) 18:359–73. doi: 10.1016/j.media.2013.12.002

34. Cuocolo R, Stanzione A, Castaldo A, De Lucia DR, Imbriaco M. Quality control and whole-gland, zonal and lesion annotations for the PROSTATEx challenge public dataset. *Eur J Radiol*. (2021) 138:109647. doi: 10.1016/j.ejrad.2021.109647

35. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv:1412.6980*. (2017).

36. Ju J, Zhang Q, Xu P, Liu T, Li C, Guan Z. PZS-Net: incorporating of frame sequence and multi-scale priors for prostate zonal segmentation in transrectal ultrasound. *Adv Intell Syst*. (2024) 7:2400302. doi: 10.1002/aisy.202400302

37. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z

38. Jia H, Xia Y, Song Y, Zhang D, Huang H, Zhang Y, et al. 3D APA-Net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MR images. *IEEE Trans Med Imaging*. (2019) 39:447–57. doi: 10.1109/TMI.2019.2928056

39. Zhu Q, Du B, Yan P. Boundary-weighted domain adaptive neural network for prostate MR image segmentation. *IEEE Trans Med Imaging*. (2019) 39:753–63. doi: 10.1109/TMI.2019.2935018

40. Wang W, Wang G, Wu X, Ding X, Cao X, Wang L, et al. Automatic segmentation of prostate magnetic resonance imaging using generative adversarial networks. *Clin Imaging*. (2021) 70:1–9. doi: 10.1016/j.clinimag.2020.10.014

41. Qian Y. ProSegNet: a new network of prostate segmentation based on MR images. *IEEE Access*. (2021) 9:106293–302. doi: 10.1109/ACCESS.2021.3096665

42. Meyer A, Chlebus G, Rak M, Schindele D, Schostak M, van Ginneken B, et al. Anisotropic 3D multi-stream CNN for accurate prostate segmentation from multi-planar MRI. *Comput Methods Programs Biomed*. (2021) 200:105821. doi: 10.1016/j.cmpb.2020.105821

43. Ocal H, Barisci N, A. novel prostate segmentation method: triple fusion model with hybrid loss. *Neural Comput Applic*. (2022) 34:13559–74. doi: 10.1007/s00521-022-07188-3

44. Jia H, Cai W, Huang H, Xia Y. Learning multi-scale synergic discriminative features for prostate image segmentation. *Pattern Recognit*. (2022) 126:108556. doi: 10.1016/j.patcog.2022.108556

45. Chen T, Zhao X, Ling Q, Gong Z, Tao B, Yin Z. An automatic prostate surgical region reconstruction method based on multilevel learning. *IEEE Trans Instrum Meas*. (2022) 71:1–12. doi: 10.1109/TIM.2022.3192862

46. Bhandary S, Kuhn D, Babaiee Z, Fechter T, Benndorf M, Zamboglou C, et al. Investigation and benchmarking of U-Nets on prostate segmentation tasks. *Comput Med Imag Graph*. (2023) 107:102241. doi: 10.1016/j.compmedimag.2023.102241

47. Ma L, Fan Q, Tian Z, Liu L, Fei B, A. novel Residual and Gated Network for prostate segmentation on MR images. *Biomed Signal Process Control*. (2024) 87:105508. doi: 10.1016/j.bspc.2023.105508

48. Yu B, Zhou Q, Yuan L, Liang H, Shcherbakov P, Zhang X. 3D medical image segmentation using the serial-parallel convolutional neural network and transformer based on cross-window self-attention. *CAAI Trans Intell Technol*. (2025) 10:12411. doi: 10.1049/cit2.12411

49. Zhong J, Liu T, Piao Y, Sun W, Lu H. ProSegDiff: prostate segmentation diffusion network based on adaptive adjustment of injection features. *IEEE Signal Process Lett*. (2025) 32:1236–40. doi: 10.1109/LSP.2025.3548422

50. Liu T, Zhang M, Liu L, Zhong J, Wang S, Piao Y, et al. Cridiff: criss-cross injection diffusion framework via generative pre-train for prostate segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (2024). p. 102–112. doi: 10.1007/978-3-031-72111-3_10

51. Yan Y, Liu R, Chen H, Zhang L, Zhang Q. CCT-Unet: a U-shaped network based on convolution coupled transformer for segmentation of peripheral and transition zones in prostate MRI. *IEEE J Biomed Health Inform*. (2023) 27:4341–51. doi: 10.1109/JBHI.2023.3289913

52. Hung ALY, Zheng H, Miao Q, Raman SS, Terzopoulos D, Sung K. CAT-Net: A cross-slice attention transformer model for prostate zonal segmentation in MRI. *IEEE Trans Med Imaging*. (2022) 42:291–303. doi: 10.1109/TMI.2022.3211764

53. Nguyen KM, Fernandez-Quilez A. Segmentation uncertainty with statistical guarantees in prostate MRI. In: *2024 32nd European Signal Processing Conference (EUSIPCO)*. IEEE (2024). p. 1636–1640. doi: 10.23919/EUSIPCO63174.2024.10715449

54. Wei C, Liu Z, Zhang Y, Fan L. Enhancing prostate cancer segmentation in bpMRI: Integrating zonal awareness into attention-guided U-Net. *Digital Health*. (2025) 11:20552076251314546. doi: 10.1177/20552076251314546

55. Nai YH, Teo BW, Tan NL, Chua KYW, Wong CK, O'Doherty S, et al. Evaluation of multimodal algorithms for the segmentation of multiparametric MRI prostate images. *Comput Math Methods Med*. (2020) 2020:8861035. doi: 10.1155/2020/8861035