Complex & Intelligent Systems (2025) 11:131
https://doi.org/10.1007/s40747-024-01751-2

ORIGINAL ARTICLE

RenalSegNet: automated segmentation of renal tumor, veins, and arteries in contrast-enhanced CT scans

Rashid Khan[1,2] · Chao Chen[1,2] · Asim Zaman[3] · Jiayi Wu[1,2] · Haixing Mai[4] · Liyilei Su[1] · Yan Kang[2,3] · Bingding Huang[1]

Received: 16 July 2024 / Accepted: 14 December 2024

Abstract

Renal carcinoma is a frequently seen cancer globally, with laparoscopic partial nephrectomy (LPN) being the primary form of treatment. Accurately identifying renal structures such as kidneys, tumors, veins, and arteries on CT scans is crucial for optimal surgical preparation and treatment. However, the automatic segmentation of these structures remains challenging due to the kidney's complex anatomy and the variability of imaging data. This study presents RenalSegNet, a novel deep-learning framework for automatically segmenting renal structure in contrast-enhanced CT images. RenalSegNet has an innovative encoder-decoder architecture, including the FlexEncoder Block for efficient multivariate feature extraction and the MedSegPath mechanism for advanced feature distribution and fusion. Evaluated on the KiPA dataset, RenalSegNet achieved remarkable performance, with an average dice score of 86.25%, IOU of 76.75%, Recall of 86.69%, Precision of 86.48%, HD of 15.78 mm, and AVD of 0.79 mm. Ablation studies confirm the critical roles of the MedSegPath and MedFuse components in achieving these results. RenalSegNet's robust performance highlights its potential for clinical applications and offers significant advances in renal cancer treatment by contributing to accurate preoperative planning and postoperative evaluation. Future improvements to model accuracy and applicability will involve integrating advanced techniques, such as unsupervised transformer-based approaches.

Keywords RenalSegNet · Kidney cancer · Contrast-enhanced CT scans · MedSegPath · Deep learning

Introduction

Kidney cancer is a major issue in global public health, as kidney diseases affect millions of people worldwide [1]. The diagnosis and management of renal diseases, including cancer, often rely on accurate imaging methods such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US) [2]. These image modalities provide clinicians detailed information and are making contributions to essential aspects of renal structure and function, which enable these technologies to help study and treat kidney diseases by revealing information about renal blood flow and tissue characteristics. Enhanced CT scans are of particular importance in clinical practice as they offer a detailed analysis of renal conditions [3]. However, image interpretation by experienced radiologists may be lengthy and error-prone, especially in complicated scenarios, such as kidney cancer or chronic kidney disease. The manual segmentation of renal tumors, veins, and arteries in acquired images is a laborious and time-consuming undertaking that leads to high inter-observer variability [4]. Medical image segmentation will be equally critical to some clinical activities such as computer-aided diagnosis (CAD), and constructing visual models [5]. With the help of a definite visual representation of renal anatomy, surgeons will be able to find renal tumor arteries, arrange the event before surgery, and then perform a much better surgery [6]. Performing surgical planning based on 2D

✉ Bingding Huang
huangbingding@sztu.edu.cn

1 College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518188, China

2 College of Applied Sciences, Shenzhen University, Shenzhen 518060, China

3 College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen, China

4 Department of Urology, Third Medical Center, Chinese PLA General Hospital, Beijing 100036, China

Published online: 07 January 2025

images is not equivalent in terms of accuracy and best surgical choices about the area that will be incised and removal of tissue during renal cancer surgery with a 3D image [7].

Recently, deep learning techniques have demonstrated promise in enabling the automation of medical image segmentation tasks [8]. These methods utilize convolutional neural networks (CNNs) to directly identify complex patterns and features from medical imaging data, which can ensure the accurate segmentation of kidney structures [9]. The primary aim of developing RenalSegNet is to overcome the limitations of existing state-of-the-art models in precisely segmenting renal structures, particularly in challenging and low-contrast areas. U-Net and its derivatives, such as UNet + + [10], R2U-Net [11], and Att-UNet [12], have been successful in incorporating more high-level semantic information. This has been achieved by creating either deeper mappings or skip connections that will help capture the deep semantic information but will frame spatial details [13]. However, recent advancements like attention mechanisms and transformer-based methods offer potential solutions to these constraints [14].

Furthermore, additional challenges in developing reliable segmentation networks for kidney-related structures include a lack of annotated training data and the interpretability of deep learning models. However, as shown through a recent study, deep learning models can potentially automate the segmentation of abdominal organs and tumors. Thus, models like FastFCN [15], Anatomynet [16], and Attention U-Net [12] encourage further research to enhance treatment outcomes for renal cancer. Despite recent research encountering challenges such as limited annotated data and interpretability problems in deep learning models, they prioritize segmenting crucial structures, including the kidney, tumors, intricate renal arteries, and low-contrast segments of renal veins. The proposed framework is designed based on superior state-of-the-art methods optimized explicitly to capture and refine high-level features in CT scans, as it includes innovations like FlexEncoder Block and MedSeg-Path Mechanism. These improvements enable RenalSegNet to more accurately segment thin and closely aligned structures, such as veins and arteries, while overcoming issues existing models face in low-contrast regions. RenalSegNet improves the Challenges of baseline models with its novel architecture, achieving better results than current state-of-the-art methods in qualitative and quantitative evaluations. This is particularly important in clinical practice, where accurate segmentation of renal structures plays a key role in preoperative planning and postoperative assessment. To address the previously mentioned challenges, The study proposes a RenalSegNet deep learning approach that enables an automatic and accurate segmenting of the different parts of the kidney cancer structure using a segmentation network. More specifically, the network can detect and outline the kidney, the kidney tumor, veins, and arteries, in contrast to enhanced CTs, which are readily available through online repositories and on the KiPA22 public dataset [17]. The main aspects of this research are as follows:

1. This paper presents a novel encoder-decoder-based framework called *RenalSegNet* designed to automatically segment renal cancer structures in contrast-enhanced CT images. In particular, RenalSegNet accurately delineates the kidney, renal tumors, veins, and arteries, demonstrating its suitability for advanced medical image analysis in oncological planning and diagnosis applications.
2. The FlexEncoder Block for RenalSegNet is designed to effectively extract features from the input CT images. This is achieved using Conv3D layers to obtain multi-scale feature information from different dimensions. FlexEncoder incorporates Group Normalization (GN), and a Squeeze-and-Excitation (SE) module is optionally added to improve the recalibration of feature maps. The decoder block (FlexUpscale Module) uses the Conv3DTranspose layer to upsample the low-resolution feature maps to generate the high-resolution segmentation maps.
3. A novel MedSeg-Path Block was introduced, placed between the encoder and decoder layers of the model. This mechanism minimizes gradient vanishing and elevates feature refinement, integration, and aggregation, thus achieving more enhanced segmentation of kidney morphology.
4. Despite being a relatively simple model, RenalSegNet improved quantitative performance across six multiple metrics and qualitatively enhanced each of the renal structures, outperforming current state-of-the-art methods. The results on the KiPA22 (Kidney PArsing Challenge 2022) dataset demonstrate network robustness and generalizability as a new benchmark in medical image segmentation.

The rest of this paper is structured as follows: In Related Work, a comprehensive review of methods, aims, limitations, and datasets from various medical image segmentation techniques, focusing on approaches used for renal structure segmentation. In the Proposed Network Architecture section, RenalSegNet, an encoder-decoder-based framework, and its key components are presented. In Experiment Configurations, the experimental setup, including the datasets, evaluation metrics, and training procedure used to assess the framework, is presented. In Results and Discussions, the performance of the proposed approach is demonstrated by presenting the outcomes of the experiments. A detailed comparison with existing methods and an in-depth discussion of the results are provided. Conclusion section summarizes the key findings and suggests directions for future work.

Related work

Medical image segmentation, particularly in the context of renal cancer treatment, has seen significant advancements driven by deep learning techniques [18, 19]. Given the recent achievement of deep convolutional neural networks (DCNN) in various vision tasks, multiple DCNN-based methods-based approaches have been proposed for kidney-related structure segmentation [20, 21]. In particular, U-Net has been a focus concerning its gradual extraction of semantic information [22]. Despite their excellent performance, they still find it challenging to extract the context information. Advanced deep learning networks, such as PSPNet [23], PoolNet [24], and DeepLabV3 [25], address this by incorporating multi-scale context information. However, these approaches often find it difficult to adapt receptive field sizes dynamically depending on the target object's size. Moreover, an effective and simple V-shape network and Dice loss were designed to perform volumetric medical image segmentation [26]. Furthermore, the data attribute and network design incorporated by instinctive knowledge have demonstrated excellent performance in kidney tumor segmentation [6]. Similarly, Refs. [27] and [28] employed a comparable strategy, utilizing neural architecture search (NAS) to directly explore network designs from data and improved Precision in abdominal organ segmentation.

Furthermore, a recent study in computational anatomical organ and tumor segmentation has focused on automating segmentation tasks for abdominal organs and tumors [29, 30]. For instance, the multi-scale U-Net was designed along with a feature recalibration strategy for renal tumor segmentation [28, 31], while the pyramid-like convolution structure has been proposed to extract inner-slice features of renal tumors [32]. Similarly, a multi-model fusion down-sampling block introduced for pancreas and tumor segmentation shows how deep learning can improve treatment outcomes in cases of abdominal cancers [33, 40]. Deep learning approaches have also impacted other medical domains. Recent work applied a multilayer perceptron for neonatal sleep–wake classification using EEG data [46]. CNN-based models have shown potential in brain tumor characterization, particularly with transfer learning for pediatric cases [47]. Additionally, GAN-based techniques have been used for synthetic medical image classification [48]. Table 1 provides an overview of previous studies on medical image segmentation using deep learning approaches, highlighting their diverse goals and specific limitations.

Most of the existing renal cancer structure segmentation models focused on kidney and tumor segmentation [23, 29, 33, 37, 38]. In this era, these methods have made great achievements in this field as they all aimed to have definitive segmentation of kidney and renal tumors, leaving the secondary structures such as renal arteries and veins deserted regarding preoperative planning of renal surgery [33, 34, 38]. Segmentation of the renal arteries and veins could be important for many intricate surgical robotic operations and the complete delivery of anatomic operative strategies [41–43, 45]. In this context, there is a need for automated segmentation of the renal cancer structures, encompassing the kidney, renal tumor, renal arteries, and renal veins, for comprehensive analysis of renal cancer. To overcome the challenges, a deep encoder decoder-based network called RenalSegNet is proposed to robustly segment renal cancer anatomical structures, including the kidney, renal tumor, and vascular structures (renal arteries and veins).

Proposed network architecture

RenalSegNet is an encoder-decoder-based framework developed to automate the segmentation of renal cancer structures from contrast-enhanced CT images. It segments key renal anatomy, including kidneys, renal tumors, veins, and arteries, essential for oncological planning and diagnosis. However, the segmentation process faces challenges such as low contrast in CT images, anatomical variations, and multi-subtype tumors, which complicate accurate delineation and require a robust model to address these issues. Figure 1 highlights these key challenges in renal segmentation.

To overcome these challenges, RenalSegNet utilizes an innovative architecture composed of three key components: the FlexEncoder Block, the MedSegPath mechanism, and the FlexUpscale module, as illustrated in Fig. 2. The proposed architecture is extensively adapted to segment the primary renal anatomy, including the kidney, renal tumors, veins, and arteries, which have great potential in oncological treatment planning and diagnosis. The FlexEncoder Block captures multi-scale feature representation by utilizing Conv3D layers to fetch features from input CT images from all dimensions efficiently. Furthermore, GN is employed in the FlexEncoder Block to instill model reliability. In addition, an optional Squeeze-and-Excitation module is added to recalibrate the feature maps and refine the feature map representation. The decoder section, FlexUpscale, utilizes Conv3DTranspose layers to up-sample the low-resolution feature maps to obtain high-resolution feature maps.

FlexEncoder Block

The FlexEncoder Block of the proposed RenalSegNet model uses several convolution layers to extract features while retaining spatial information from input CT images (as shown in Fig. 3, below). The residual block structure uses an intermediate feature modification strategy, where there exists a sequence of convolution layers that produce feature maps with different numbers of filters.

Table 1 Summary of Methods, Aims, Limitations, and Datasets of Various Medical Image Segmentation Techniques

Authors/Refeences	Methods	Aims	Limitations of the Methods	Datasets
Isensee et al. [34]	A 3D U-Net with a residual structure	Solving the problem of gradient vanishing and winning the KiTS2019 challenge	Lacking a thorough examination of distinctions between the standard U-Net and its variants	KiTS2019 dataset
Meyer et al. [35]	An anisotropic 3D multi-stream CNN architecture	The proposed network design fuses information from anisotropic images alleviating the need for image resampling to isotropic voxel size and is computationally less expensive	Addressing automated detection methods for such outliers and assess the time needed to achieve clinically acceptable segmentations through correction	ProstateX dataset In-House dataset
Dong et al. [36]	A novel mesh network (MNet)	It combines 2D and 3D CNN to balance spatial information for anisotropic medical image segmentation	Leading to increased computational complexity and memory requirements compared to standalone 2D or 3DCNNs	LiTs 2017 KiTs2019 BraTs2020 PROMISE2012
Feng et al. [37]	Context Pyramid Fusion Network (CPFNet)	Addressing the existing U-shaped CNN that lack sufficient context information extraction capabilities	Retinal linear lesion dataset collects from 38 subjects, which may limit the generalizability of the results to larger, more diverse populations or clinical setting	ISIC2018 dataset ICGA images SegTHOR AI challenge 2018 for automated segmentation of retinal edema lesions
He et al. [38]	Meta Grayscale Adaptive Network (MGANet)	Addressing low contrast due to narrow grayscale distribution and inter-image distribution variation	Reliance on a specific dataset with limited generalizability across different imaging protocols or scanners	A abdominal CTA images from Jiangsu Province Hospital
Mou et al. [39]	Curvilinear Structure Segmentation Network (CS^2-Net)	Aims to enhance the segmentation of curvilinear structures by incorporating spatial and channel attention mechanisms	The method struggles to accurately identify curvilinear structures resembling cells, and its computational requirements for 3D segmentation are high	DRIVE, STARE, IOSTAR, CORN-1, OCTA, OCT-RPE, MIDAS, Synthetic, VascuSynth
Jin et al. [40]	Multi-Branch Segmentation Network (MBSNet)	Aims to capture and fuse both local and global information by using multi-branch	Computation complexity is notably larger than some other comparative networks	ISIC2018 Kvasir BUSI, COVID-19, LGG

The FlexEncoder Block, a cornerstone of the RenalSeg-Net architecture, is inspired by the computational efficiency principles of the Squeeze-and-Excitation (SE) mechanism. It commences with a 3D convolution operation represented as in Eq. (1),

$$O_{i,j,k} = \sum_{l,m,n} I_{i-l,j-m,k-n} \times F_{l,m,n} \tag{1}$$

where $O_{i,j,k}$ denotes the value at position (i, j, k) in the output tensor O. The convolution layer efficiently processes the width, height, and depth dimensions of input data, which is essential for medical imaging tasks. The Global Average Pooling (GAP) operator subsequently aggregates across all the spatial dimensions in Eq. (2),

$$S_i = \frac{1}{H \times W \times D} \sum_{i,j,k} O_i, j, k \tag{2}$$

where S_i represents the aggregated feature for the i^{th} channel.

Subsequently, the SE mechanism in the RenalSegNet architecture enhances feature representation and computational efficiency through two 1×1x1 convolution layers after the GAP. Let the output tensor Z after the GAP, and the

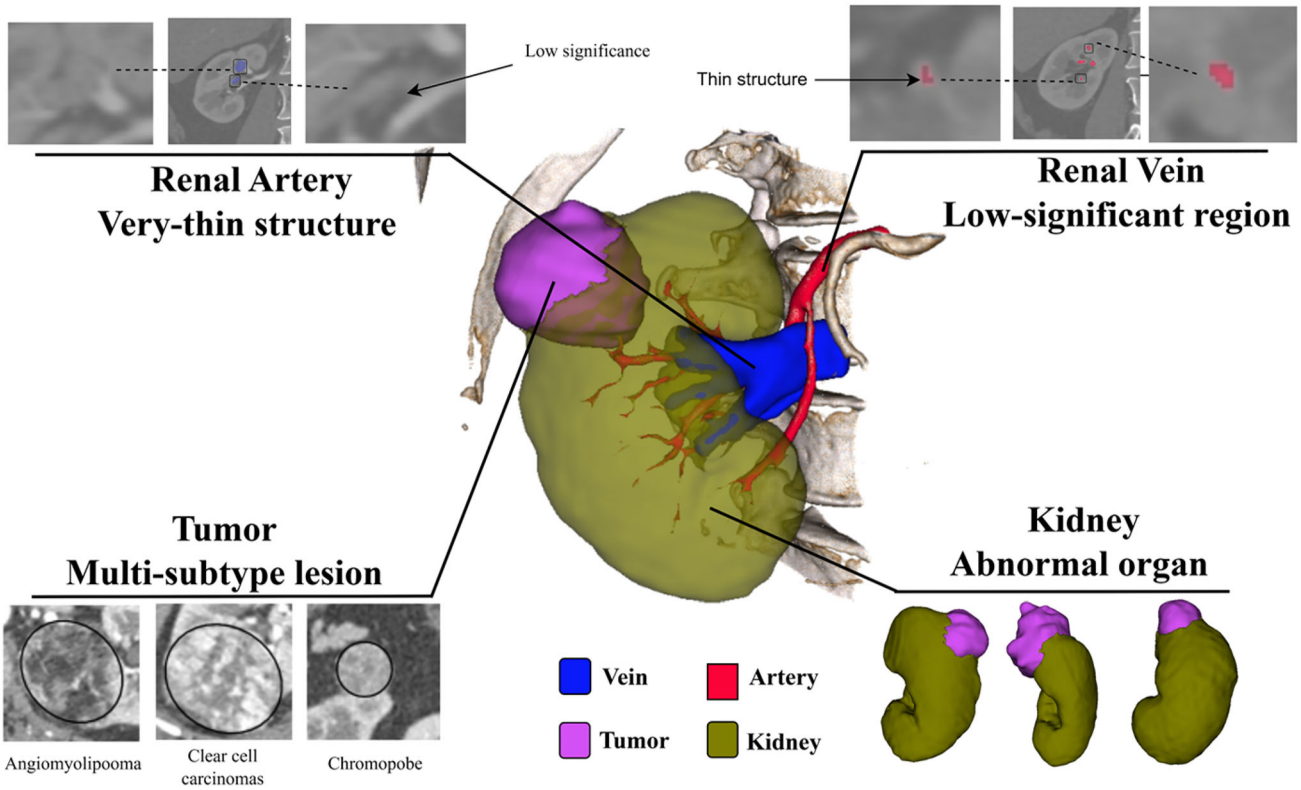

Fig. 1 Visual representation of the intricate nature of segmenting kidney-related structures and highlights the challenges ((Low contrast in the original image, aboriginal organ, low significant region, and multi-subtype tumor) in accurately delineating 3D kidneys, renal tumors, arteries, and veins

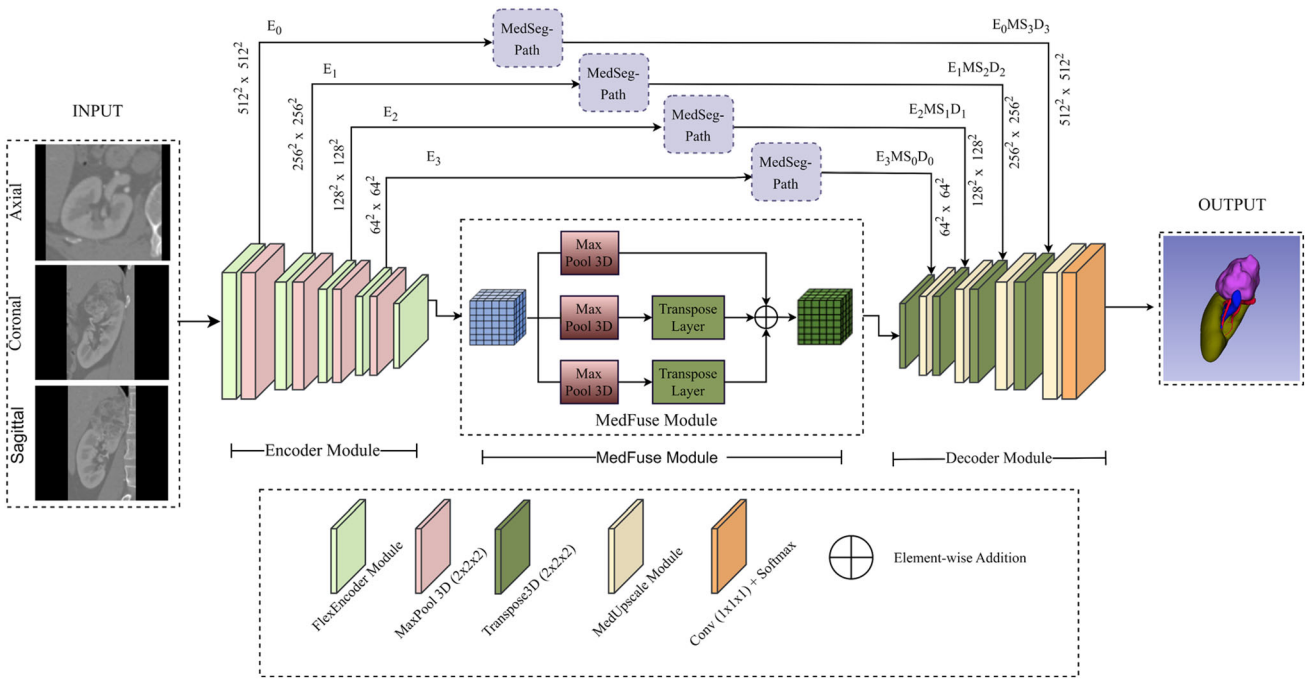

Fig. 2 RenalSegNet's network architecture consists of the FlexEncoder Block, the MedSegPath Mechanism, the MedFuse link, and the FlexUpscale Module, which are designed to automate the segmentation of renal cancer structures in the input contrast-enhanced CT images

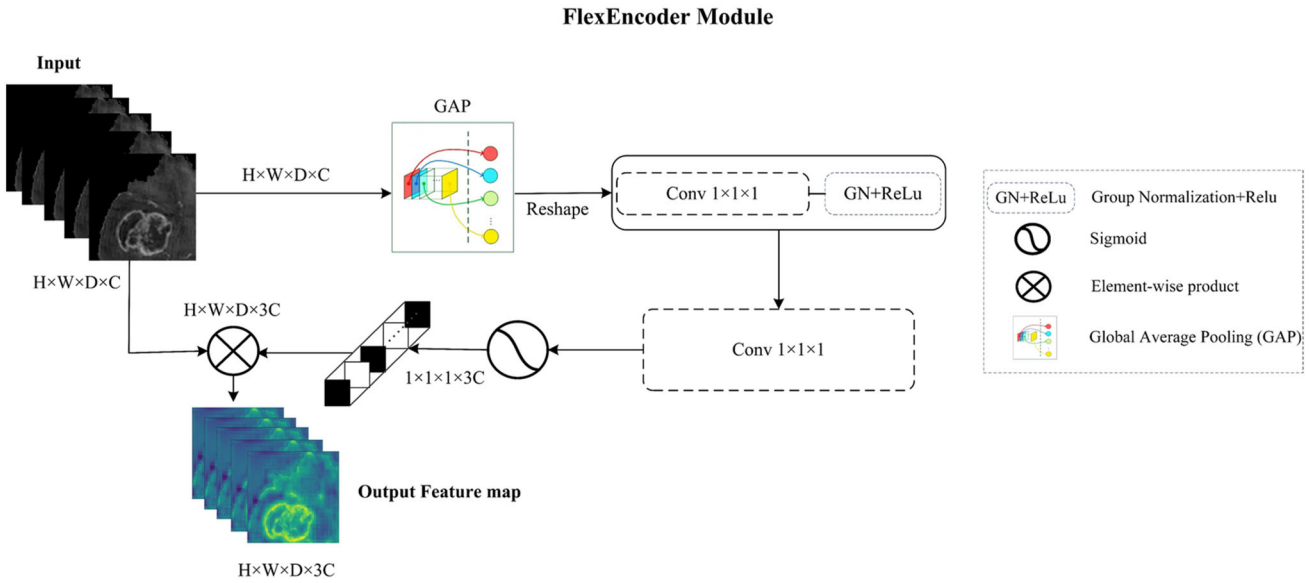

Fig. 3 FlexEncoder Block of the RenalSegNet

first channel-wise transformation is performed through two convolution layers $Conv_{1\times1\times1}(1)$ and $Conv_{1\times1\times1}(2)$ with their weight tensors $Wconv_{1\times1\times1}(1)$ and $Wconv_{1\times1\times1}(2)$ respectively. Equations (3) and (4) are the outputs tensors A and B.

$$A_{c'} = \sum_{c=1}^{C} Z_c \times W_{conv_{1\times1\times1}(1)}(c, c') \tag{3}$$

$$B_{c'} = \sum_{c=1}^{C'} A_{c'} \times W_{conv_{1\times1\times1}(2)}(c', c) \tag{4}$$

where C is the number of channels in the input tensor Z and $C\prime$ is the number of channels in the output tensor A as well as the number of filters in the first 1x1x1 convolution layer, whereas c' denotes the channel index of the output tensors A and B.

MedSegPath Mechanism

The MedSegPath mechanism, an extension of the skip connection strategy, introduces adaptive feature accumulation and feature map fusion to improve the interaction of encoder and decoder networks. This skip connection's improved connectivity approaches are based on the success of approaches such as ResPath and DenseSkip-1 [44]. Figure 4 depicts the segmentation path's visualization of the skip connection in RenalSegNet, associated with the skip connection path.

The MedSegPath block operates on an input feature map represented as X, and multiple parallel transforms are applied to capture more than one aspect of the input. Concretely, X is first convoluted with a 1x1x1 filter W_1, shown as a transformed tensor X' as in Eq. (5):

$$X'_{i,j,k,l} = \sum_{c=1}^{C} X_{i,j,k,c} \cdot W_{1,l,c} \tag{5}$$

i, j and k represent spatial locations in the 3D feature maps, l is a feature channel at each spatial location, C is the number of channels in the input feature map X, and W_1, l, c are coefficients of the convolution filter, where X is convolved with filters 3x5x5 and 3x3x3, respectively, leading to a tensor of Y,

$$Y_{i,j,k,l} = \sum_{m=-1}^{1} \sum_{n=-1}^{1} \sum_{p=-1}^{1} X'_{i+m,j+n,k+p,l} \cdot W_{2,m,n,p} \tag{6}$$

with m, n, and p representing the indices over the spatial support of the 3D convolution kernel taking values between -1 and 1.

Convolution is followed by batch-normalization at each layer, which normalizes tensors act and con into X_{norm} and Y_{norm}. After normalization, tensors receive an element-wise activation function, in this case ReLU, to yield tensors X_{act} and Y_{act}.

This series of steps occurs for N times and sums the outputs together to produce the final output tensor O and (n) is the number of iterations, as described in Eq. (7).

$$O_{i,j,k,l} = \sum_{n=1}^{N} Z^{(n)}_{i,j,k,l} \tag{7}$$

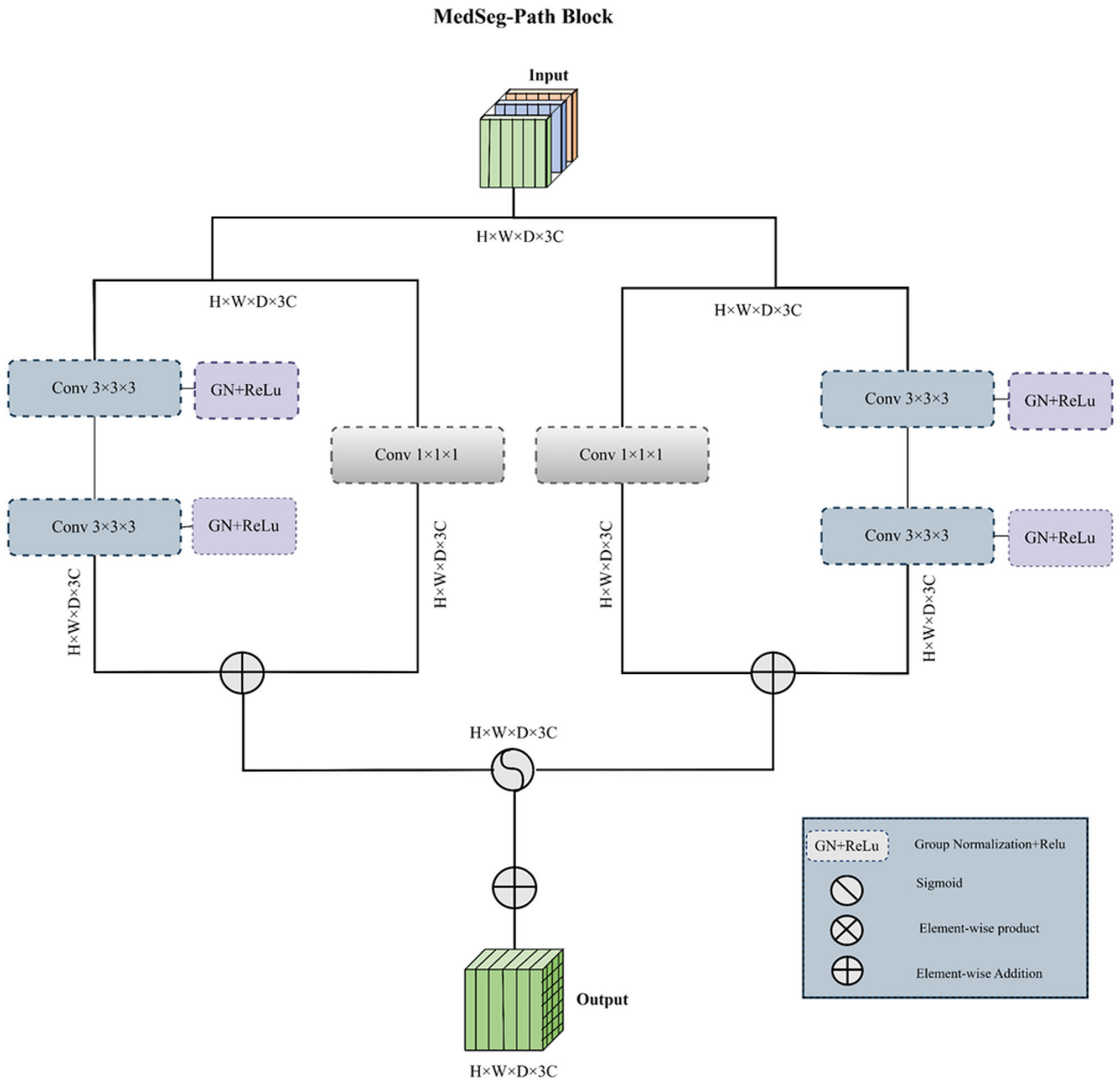

Fig. 4 Visualization of the MedSeg-Path Mechanism in RenalSegNet, showcasing the skip connection path that preserves spatial information and enhances feature propagation during the segmentation process

Additionally, the accumulation of adaptive features enriches the information in the final output tensor, which also permits capturing more complex patterns and variances prevailing in the input data. Hence, the above procedures of adaptive feature accumulation boost the formwork to retain more valuable features.

MedFuse Block in RenalSegNet

The integration of the MedFuse block in RenalSegNet involves feature-fusing both the encoder and decoder paths. It was developed to refine and fuse multi-scale contextual information that will consolidate segmentation performance. Figure 5 demonstrates this mechanism by depicting feature integration and refinement of the MedFuse block.

The input feature map contains the features from the encoder pathway and Y is the output feature map after MedFuse.

$$X_{fuse} = \sum_{c=1}^{C} X \cdot W_{fuse} \tag{8}$$

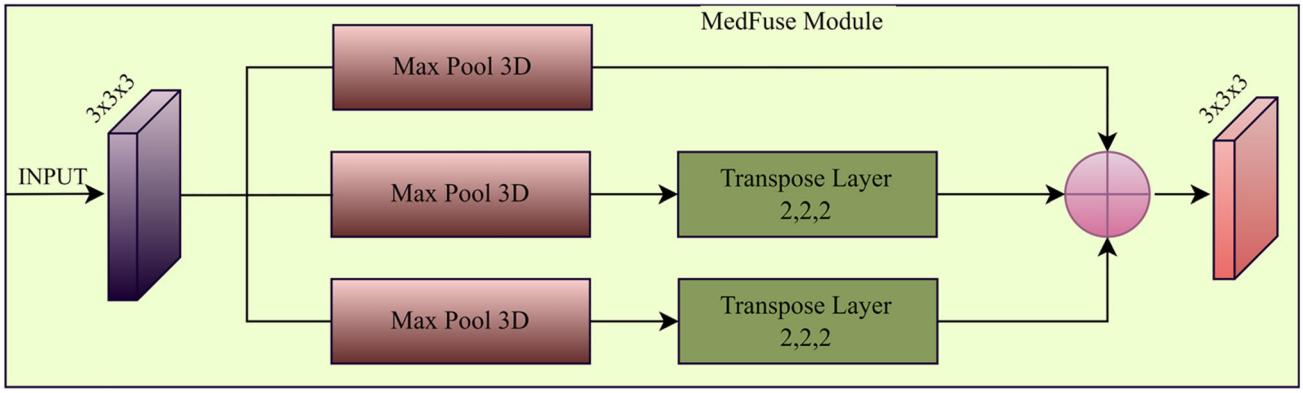

Fig. 5 Illustration of the MedFuse Block in RenalSegNet

The block applies a $1 \times 1 \times 1$ convolution operation with learnable filters W_{fuse} to yield the tensor X_{fuse} .C is the number of channels in X, and W_{fuse} denotes the convolution filter coefficients. After the convolution operation, the resulting feature map X_{fuse} is normalized through batch normalization to get the normalized feature map X_{norm} as the computer as in Eq. (9),

$$X_{\text{norm}} = \frac{X_{\text{fuse}} - \mu}{\sigma +} \tag{9}$$

where μ is the mean of X_{fuse} , σ is the standard deviation of X_{fuse} , and ϵ is a small constant to avoid division by zero using Eq. (10).

$$Y = ReLU(X_{\text{norm}}) \tag{10}$$

Then, element-wise, the ReLU activation function is applied, resulting in the final output feature map.

Decoder (FlexUpscale Module) in RenalSegNet

The FlexUpscale module in RenalSegNet is designed to upsample encoded features and reconstruct the segmentation map to the original input resolution. This process involves a series of transposed convolutions (deconvolutions), skip connections, and convolutional layers that gradually increase the spatial resolution of the feature maps while integrating fine-grained details from the encoder.

The decoder starts by applying transposed convolutions to the encoded feature map. If X is the input feature map and K is the kernel, the transposed convolution operation is given by Eq. (11):

$$Y(i,\ j,\ k) = \sum_{m} \sum_{n} \sum_{p} X(m,\ n,\ p) K(i - m,\ j - n,\ k - p) \tag{11}$$

where, Y is the output feature map $(i,\ j,\ k)$, and $(m,\ n,\ p)$ are the spatial coordinates of the output and input feature maps. This operation effectively increases the spatial dimensions of the feature map by a factor corresponding to the stride of the transposed convolution. To retain spatial information and details lost during down-sampling, the decoder incorporates skip connections from the encoder layers using the SegPath as in Eq. (12),

$$S_{\text{out}} = ReLU(X + G(X)) \tag{12}$$

where $G(X)$ is the output of two consecutive 3D convolutional layers applied to the input X. Group Normalization (GN) and ReLU activation follow each convolutional layer within. Adding the original input X to the output G forms a residual connection that aids in preserving spatial details. In Eq. (13), each Decoder block in the FlexUpscale Module applies two sets of 3D convolutions, and dropout is used to prevent overfitting, enhancing the robustness of the model.

$$Z = ReLU(GN(K \times X)) \tag{13}$$

The decoder's final output is a high-resolution segmentation map generated by applying a final convolutional layer with a softmax activation function as in Eq. (14),

$$O = softmax(K_{\text{f}} \times X) \tag{14}$$

where, K_{f} is the kernel of the final convolutional layer, and O is the output segmentation map. That generates the segmentation map with softmax activation, providing a detailed and accurate probability map for each class.

Experiments configurations

Dataset

In the experiment, the publicly accessible KiPA22 dataset [17], which includes 70 CT scans, was used. The dataset comprises 13,846 slices, with a mean spacing of (0.6328, 0.6328, 0.6328) and a mean size of (150, 150, 198). The minimum volume of the renal vein is merely 1.6 cubic centimeters, whereas that of the renal artery is only 0.9 cubic centimeters. The dataset was split into three sets (8:1:1 ratio): training (80%), validation (10%), and testing (10%). To improve the robustness of the model and avoid overfitting, several data augmentation techniques, including mirroring, flipping, rotating, scaling, and Gaussian noise, were applied to the training set. For each image, ten images with augmented data were generated.

Implementation details

The model was trained for 150 epochs, with 297 iterations in each epoch. The batch size was set to 2, and the patch size was cropped to $128 \times 128 \times 128$. The model was trained with Adam with a learning rate of 1e-4. The training loss function combined dice loss and focal loss. During the training process, the loss, accuracy, IOU score, precision, recall, and dice score were recorded for both the training and validation sets, and these metrics were visualized to aid in understanding the training process. RenalSegNet was implemented using the TensorFlow framework and trained on an NVIDIA A100 GPU.

Evaluation metric

The proposed segmentation approach's performance was evaluated using area and distance-based metrics. Six evaluation metrics were utilized: Dice, Hausdorff distance (HD), IOU, Precision, Average Hausdorff Distance (AVD), and Recall. The dice Similarity Coefficient (DSC) evaluates the area-based overlap index. AVD evaluates the surface's coincidence for stability and less sensitivity to outliers. Hausdorff Distance, which is sensitive to outliers, is also used to compare the segmentation quality of outliers further. For each target in one image, the ground truth segmentation point set is denoted as X, and the predicted set as Y. During the validation phase, additional metrics such as loss, accuracy, and IOU were continuously monitored, with Precision, recall, and Dice score calculated at each epoch to fine-tune hyperparameters. A combination of Dice loss and focal loss was applied during training, improving the model's robustness in handling class imbalances. Validation performance was tracked using loss curves across all key metrics, as depicted in Fig. 6.

DSC (Dice Similarity Coefficient)

In Eq. (15,16), Dice is widely used to evaluate the overlap between the predicted segmentation and the ground truth labels. It is defined as twice the area of the overlap between the two sets, divided by the total number of pixels in both sets:

$$\text{Dice} = \frac{2 \times \text{ Area of overlap}}{\text{Total area}} \tag{15}$$

$$DSC(X, Y) = \frac{2|X \cap Y|}{|X|+|Y|} \tag{16}$$

where X and Y represent the predicted and ground truth sets, respectively.

Hausdorff Distance (HD)

The Hausdorff Distance (HD) between two finite point sets X and Y is defined by Eq. (17)

$$HD(X, Y) = \max(p(X, Y), p(Y, X)) \tag{17}$$

where $p(Y, X)$ is called the directed Hausdorff distance and given in Eq. (18),

$$p(X, Y) = \max_{a \in X} \min_{b \in Y} \|a - b\| \tag{18}$$

$\|a - b\|$ is Euclidean distance.

AVD (Average Hausdorff Distance)

The Average Hausdorff Distance is the HD average across all points. The AVD is recognized for its stability and lower susceptibility to outliers than the HD. It is formulated by Eq (19),

$$AVD(X, Y) = \max(d(X, Y), d(Y, X)) \tag{19}$$

where, $d(X, Y)$ is the directed Average Hausdorff distance that is given in Eq. (20),

$$d(X, Y) = \frac{1}{N} \sum_{a \in X} \min_{b \in Y} \|a - b\| \tag{20}$$

Recall

The recall metric reflects the RenalSegNet model's capacity to correctly identify all actual positive instances. It is the ratio of true positives to the sum of true positives and false negatives is formulated in Eq. (21),

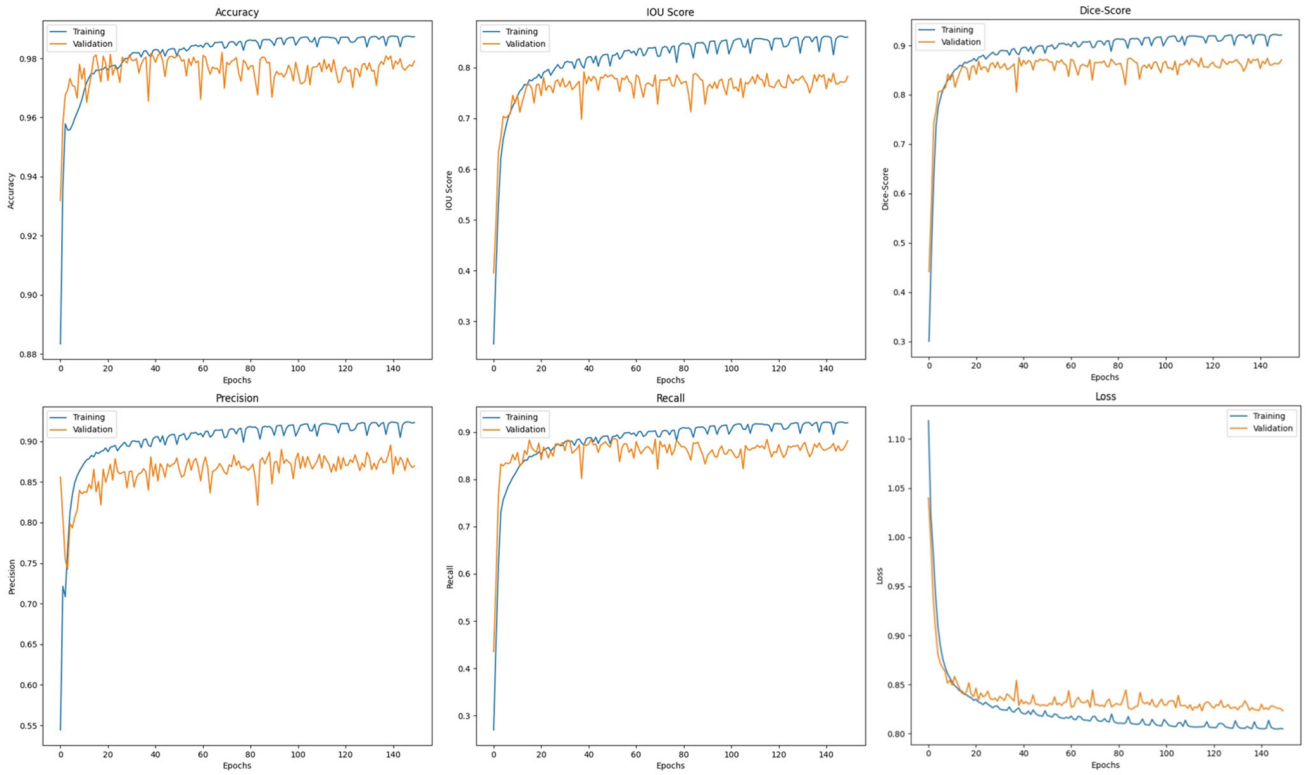

Fig. 6 Training and validation loss curves for various evaluation metrics (accuracy, IOU, Dice, Precision, Recall, and loss) in RenalSegNet

$$\text{Recall} = \frac{TP}{TP + FN} \tag{21}$$

Precision

Precision Eq. (22) indicates RenalSegNet's ability to correctly identify true positive instances. It is defined as the ratio of true positives to the sum of true positives and false positives:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{22}$$

IOU

The Intersection-over-Union (IoU), Eqs. (23–24), is the intersection area over the union of the predicted segmentation and the ground truth,

$$\text{IoU} = \frac{\text{Area of overlap}}{\text{Area of union}} \tag{23}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{24}$$

The IOU score helps measure the accuracy in RenalSegNet evaluation, and higher IOU scores generally indicate better performance. Figure 6 shows the training and validation loss curves for various evaluation metrics network trained using the proposed training strategy.

Algorithm

Algorithm 1 depicts the pseudo-code of the presented RenalSegNet.

Input: CT Images X

Output: Segmentation masks Y

Pre-process X and apply data augmentation DA to enlarge training dataset: $X = \text{DA}(\text{Pre-Process}(X))$;
Set $x^0_{\text{Encoder}} = X$;
for i = 1 to 4 **do**
    Perform FlexEncoder module: $x^i_{\text{FlexEncoder}} = \text{FlexEncoder}(x^0_{\text{Encoder}})$;
    Using MedSegPath module to process the feature maps passed through skip connection: $\overline{x}^i_{\text{MedSegPath}} = \text{MedSegPath}(x^i_{\text{FlexEncoder}})$;
    Pooling operation: $x^0_{\text{Encoder}} = \text{MaxPool3D}(x^i_{\text{FlexEncoder}})$;
end for
Apply MedFuse module: $x^0_{\text{Decoder}} = \text{MedFuse}(x^0_{\text{Encoder}})$;
for i = 4 to 1 **do**
    Apply Transpose3D operation: $\overline{x}^i_{\text{Decoder}} = \text{Transpose3D}(x^0_{\text{Encoder}})$;
    Feature fusion operation: $x^i_{\text{Decoder}} = \text{Concat}(\overline{x}^i_{\text{Decoder}}, \overline{x}^i_{\text{MedSegPath}})$;
    Apply MedUpscale Module: $x^0_{\text{Decoder}} = \text{MedUpscale}(x^i_{\text{Decoder}})$;
end for
Apply convolution and softmax to generate the final segmentation map: $Y = \text{Softmax}(\text{Conv}(x^0_{\text{Decoder}}))$;
Return Y;

Algorithm 1 Pseudo-code for RenalSegNet

Results and discussions

The proposed RenalSegNet accomplishes precise 3D segmentation of renal structures by connecting different representations and providing detailed feature representation. The comparative studies' quantitative and qualitative evaluations in comparative studies show that RenalSegNet outperforms other state-of-the-art models with high Dice, HD, AVD, Recall, Precision, and IoU scores for tumors, veins, and arteries. The ablation study provides a deeper understanding of RenalSegNet's success, revealing significant contributions from the FlexEncoder, MedSegPath, and MedUpscale (decoder) modules. It also highlights the importance of tuning hyperparameters, such as filter numbers and dropout rates, to optimize performance, further solidifying the effectiveness of the proposed approach.

Quantitative analysis of metric superiority

Tables 2, 3, 4, 5 and 6 present the quantitative results for the various structures of kidney segmentation in four parts: kidney, tumor, artery, and vein. To demonstrate the superiority of the RenalSegNet framework, a comparison was made with three state-of-the-art (SOTA) medical image segmentation methods, including DensebiasNet [45], Mnet [36], CS2-Net [39], and MBSNet-3D [40], along with U-Net [22] as a baseline. The metrics used for evaluation include Dice (%), IOU (%), Recall (%), Precision (%), HD (mm), and AVD (mm). These six metrics comprehensively represent segmentation performance, capturing both the segmented regions' accuracy and quality. The proposed approach was thoroughly evaluated on the test set, and statistical validation using paired t-tests confirmed the significance of its improvements in Dice scores. The results show that RenalSegNet outperforms baseline methods, with p-values < 0.05 for Dice scores in kidney, tumor, artery, and vein segmentation tasks, highlighting the robustness and reliability of its performance.

The RenalSegNet performs significantly in most metrics.

Table 3 shows RenalSegNet outperformed other models with a Dice score of 77.45% compared to U-Net (72.40%), DensebiasNet (75.89%), Mnet (72.79%), CS2-Net (73.49%), and MBSNet-3D (75.34%). It achieved the highest IOU (63.54%) and Recall (86.19%) coupled with a competitive precision (70.38%). Also, RenalSegNet had the lowest Hausdorff Distance of all others, scoring at 14.64 mm, and the lowest Average Volume Difference at 0.53 mm, thus showing its superior performance for artery segmentation.

The proposed framework outperforms other state-of-the-art methods in all renal structures in average cases of renal segmentation. RenalSegNet model achieved the highest Dice score of 86.25%, surpassing DensebiasNet [40] (84.63%), Mnet [36] (81.93%), CS2-Net [39] (81.78%), MBSNet-3D [40] (83.61%), and U-Net [22] (83.61%). RenalSegNet achieved the highest IOU at 76.75% and had the top recall rate of 86.69%. The suggested approach shows significant outcomes with an accuracy of 86.48%, an HD of 15.78 mm, and an AVD of 0.79 mm, proving its reliability and Precision in 3D renal structure segmentation. Figure 7 shows six different metrics of comparative results to evaluate SOTA methods better. RenalSegNet achieves excellent performance in most cases using the KiPA dataset. It demonstrates that the proposed network is an advanced level of accurate segmentation of renal structures in contrast-enhanced CT scans.

Qualitative analysis of RenalSegNet on KiPA dataset

The KiPA dataset was used for visual analysis to evaluate renal structure segmentation accuracy. The proposed model captures the complex morphology of the veins, arteries, kidneys, and tumors. The findings indicate that the proposed approach produces smoother and more continuous segmentations, particularly in challenging areas where SOTA models often generate fragmented or imprecise boundaries. The proposed RenalSegNet has exceptional visual superiority and offers guidance in renal surgical supervision, as shown in Fig. 8. The smoothness of the surface in segmentation results is more than that of the ground truth according to visualization results.

In Fig. 8, a small isolated false positive area was to be segmented in the renal vein in the first case. However, a false

Table 2 Quantitative Evaluation of RenalSegNet for Vein Segmentation, Performance comparison of RenalSegNet with other SOTA models (U-Net, DensebiasNet, Mnet, CS2-Net, MBSNet-3D) using metrics Dice (%) IOU (%), Recall (%), Precision (%), HD (mm) and AVD (mm)

Methods	Vein						
	Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)	P-Value
U-Net [22]	77.22	63.15	73.70	82.56	17.15	0.97	0.0054
DensebiasNet [45]	78.90	65.50	76.59	82.42	19.39	1.04	0.0088
Mnet [36]	75.48	60.97	71.07	81.81	29.35	1.08	0.0117
CS2-Net [39]	74.98	60.44	69.97	82.30	24.68	1.11	0.0149
MBSNet-3D [40]	75.20	61.20	69.60	83.08	18.67	1.11	0.0284
RenalSegNet(our)	79.99	66.65	75.31	85.39	19.07	0.90	–

Table 3 Quantitative evaluation of RenalSegNet for artery segmentation on KiPA Dataset

Methods	Artery						
	Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)	P-Value
U-Net [22]	72.40	57.09	82.70	64.83	23.21	0.88	0.0972
DensebiasNet [45]	75.89	61.70	83.73	69.75	21.66	0.92	0.0861
Mnet [36]	72.79	57.79	80.53	67.29	20.54	0.76	0.0079
CS2-Net [39]	73.49	58.83	71.61	76.27	15.2	0.67	0.0173
MBSNet-3D [40]	75.34	60.95	78.86	78.86	15.25	0.67	0.2100
RenalSegNet(our)	77.45	63.54	86.19	70.38	14.64	0.53	–

Table 4 Quantitative evaluation of kidney segmentation Performed by RenalSegNet on KiPA Dataset

Methods	Kidney						
	Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)	P-Value
U-Net [22]	95.86	92.06	96.11	95.63	13.05	0.64	0.0663
DensebiasNet [45]	95.70	91.80	95.62	95.83	20.99	0.62	0.0644
Mnet [36]	95.30	91.04	95.60	95.06	16.22	0.67	0.0045
CS2-Net [39]	95.50	91.51	95.65	95.50	15.51	0.65	0.0110
MBSNet-3D [40]	95.90	92.13	95.91	95.91	16.27	0.61	0.0222
RenalSegNet(our)	95.89	92.12	94.66	97.17	14.85	0.59	–

Table 5 Quantitative analysis of tumor segmentation performance using the suggested model on the KiPA Dataset using six different metrics and comparing with four different SOTA approaches (U-Net, DensebiasNet, Mnet, CS2-Net, MBSNet-3D)

Methods	Tumor						
	Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)	P-Value
U-Net [22]	88.95	80.58	83.40	95.91	17.25	1.16	0.0395
DensebiasNet [45]	88.01	79.01	83.65	93.86	21.54	1.32	0.0203
Mnet [36]	84.13	73.03	77.32	93.54	34.35	2.24	0.0011
CS2-Net [39]	83.16	72.36	76.45	92.36	15.47	1.759	0.0003
MBSNet-3D [40]	87.98	79.38	82.46	82.46	10.71	1.25	0.0571
RenalSegNet(our)	91.68	84.69	90.60	92.97	14.54	1.15	–

Table 6 Quantitative assessment of average segmentation performance across all renal structures using the proposed network on the KiPA Dataset

Methods	Average						
	Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)	P-Value
U-Net [22]	83.61	73.22	83.98	84.73	17.67	0.91	0.0287
DensebiasNet [45]	84.63	74.50	84.90	85.47	20.90	0.97	0.0074
Mnet [36]	81.93	70.71	81.13	84.43	25.12	1.19	0.0020
CS2-Net [39]	81.78	70.78	78.42	86.61	17.72	1.05	0.0014
MBSNet-3D [40]	83.61	73.42	81.71	85.08	15.23	0.91	0.0184
RenalSegNet(our)	86.25	76.75	86.69	86.48	15.78	0.79	–

positive area connected with a proper renal artery was segmented in the second case. In the third and fourth cases, the method proposed performed close to ground truth in segmentation. The fourth case had significant renal tumor extremity segmentation.

In Fig. 9, the qualitative evaluation demonstrates the visual superiority of the proposed framework compared to five state-of-the-art models. The blue, green, red, and purple regions represent veins, kidneys, arteries, and tumors. Rows 1 and 2 depict the veins within the kidney structure, while Rows 3 and 4 show the arteries. In contrast, U-Net, DenseBiasNet, CS2-Net, and MBSNet have serious mis-segmentations in green boxes between the veins and the arteries. This issue is pronounced in the narrow renal hilum region, where the proximity of veins and arteries leads to significant mis-segmentation in these algorithms. The visual representation of the final out of the suggested approach closely resembles the actual vascular structure, meaning there are fewer false negatives and positives than in other models. Regarding segmenting arteries, the proposed model accurately identifies the walls of these kidneys' veins while maintaining consistency in tracking and minimizing classification errors.

Figure 9 qualitatively studies and compares the proposed model to the state-of-the-art baseline deep neural networks. As demonstrated in Fig. 10, it gives reasonable visual assistance for renal operations. This suggested deep learning-based method performs very well in recognizing organ borders and splitting the kidney and tumor while retaining the integrity of the renal parenchyma.

This suggested deep learning-based method performs very well in recognizing organ borders and splitting the kidney and tumor while retaining the integrity of the renal parenchyma. RenalSegNet provides exact tumor borders essential for determining the tumor's size, location, and extent while effectively separating the tumor from neighboring healthy kidney tissue. Accurate segmentation is crucial in preparing surgical procedures and other treatment interventions. The qualitative assessment shows that RenalSegNet generates precise segmentations that closely match expert annotations, handling the complexity and variability of renal structures (veins, arteries, and kidneys).

The Supplementary Materials (Figs. S1–S4) provide additional visualizations to further illustrate RenalSegNet's segmentation accuracy and performance. These figures provide enhanced views of the segmentation results for the veins, arteries, tumors, and whole kidneys. The improved visualizations demonstrate the model's ability to delineate complex and low-contrast structures.

Ablation experiments

The designed algorithm's ablation study, as presented in Table 7, highlights the significant performance improvements attained by combining the MedSegPath (MSP) and MedFuse (MF) components. The suggested network improves standard renal structural measures when both modules are used. Its average Dice score is 86.25%, IOU is 76.75%, Recall is 86.69%, Precision is 86.48%, HD is 15.78 mm, and AVD is 0.79 mm. Performance has a slight impact when MSP is excluded, while performance is severely impacted when both modules are excluded. MSP and MF frequently produce the highest possible accuracy and lowest error rates for vein, artery, kidney, and tumor segmentation, emphasizing their essential function in identifying complex renal structures and detecting fine-grained patterns.

This study shows that MSP and MF are necessary to achieve the optimum segmentation performance in RenalSegNet. Furthermore, the plotted graphs for each parameter provide a visual comparison of the effects of MSP and MF on RenalSegNet's performance, as shown in Fig. 11.

Analysis of segmentation challenges in RenalSegNet

Despite the overall effectiveness of the segmentation model, several limitations were identified. First, the framework struggles with accurately segmenting venous branches near the main trunk, as shown in the first row in Fig. 12. The proximity of structures makes differentiation challenging, mainly because renal veins contain low-contrast areas due

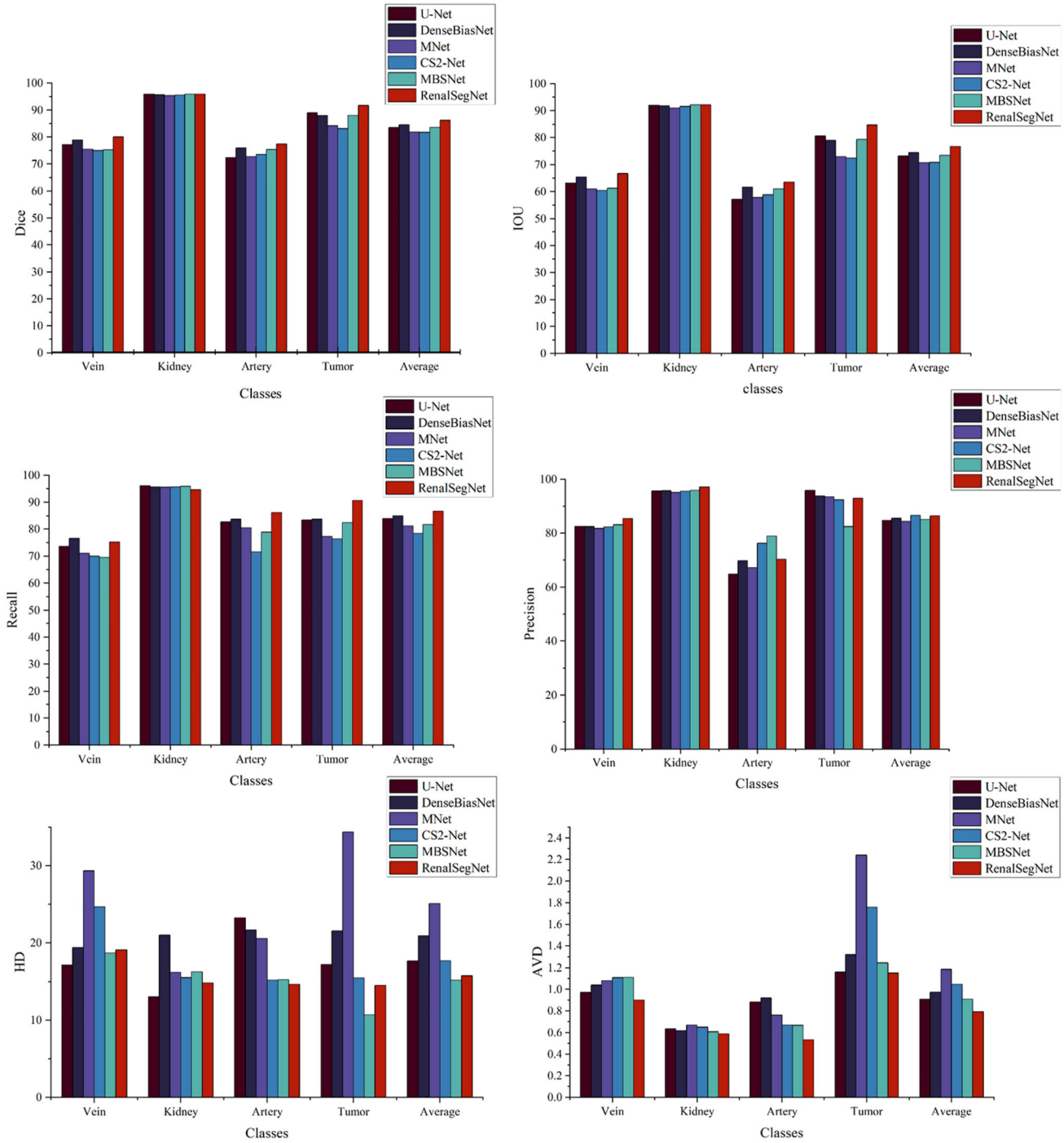

Fig. 7 Quantitative evaluation showcases RSNet's superior segmentation performance compared to other methods in contrast-enhanced CT scans (KiPA dataset) on different evaluation metrics

to minimal contrast agents, making them difficult to distinguish from the background. Second, the model tends to over-segment arterial structures, especially at their distal ends. The small size of renal arteries, which comprise only 0.27% of the image, contributes to this issue, leading to over-segmentation. A similar issue arises with tumor regions, where areas outside the tumor are incorrectly labeled. This problem may stem from the high variability in tumor appearance, as the dataset includes five different subtypes, which presents a significant challenge in managing distribution variation. Addressing these limitations will require improvements in model architecture to handle class imbalance,

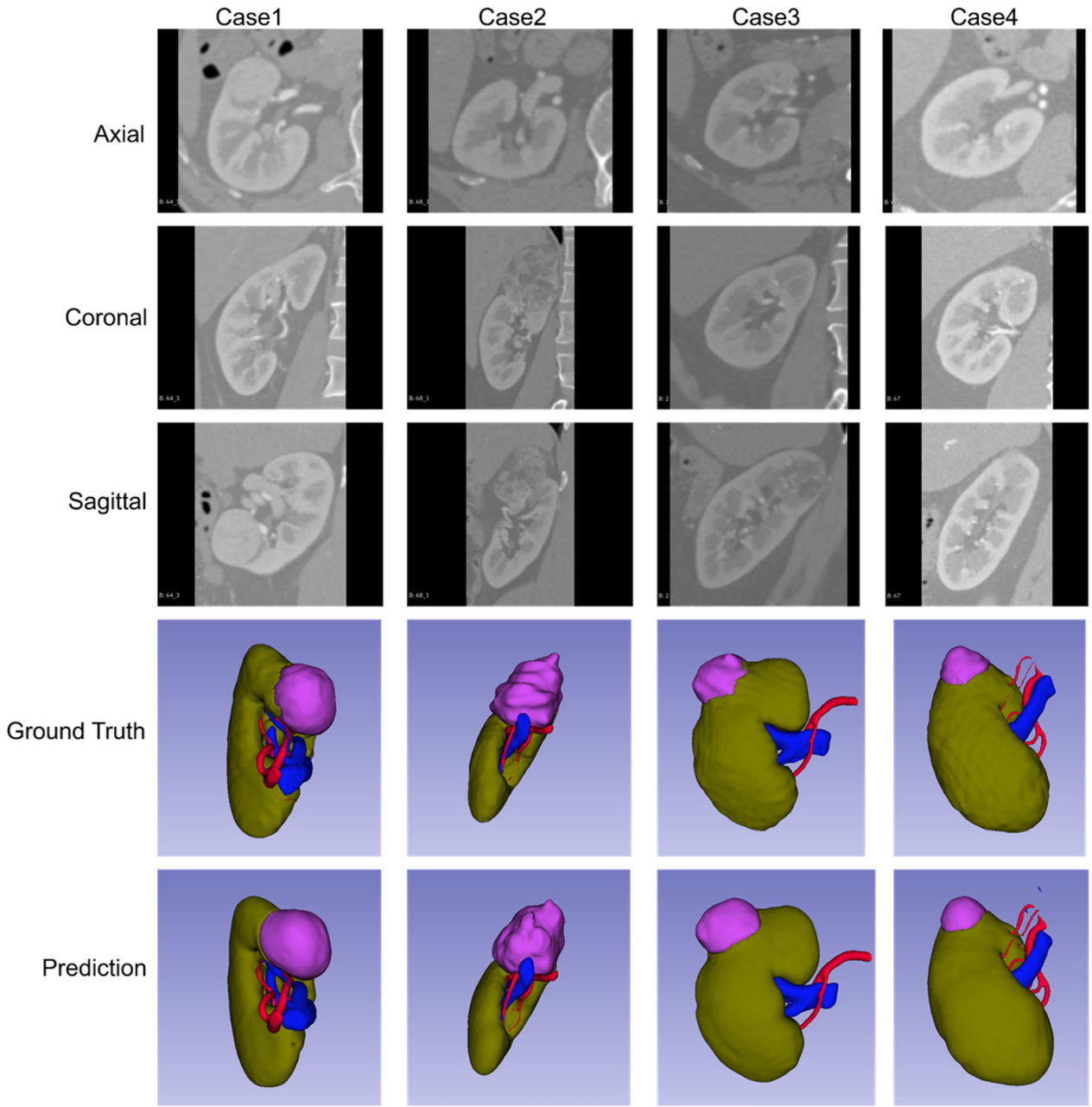

Fig. 8 The visualization of renal structure and cancer cases and their corresponding ground truths in the KiPA Dataset shows the blue, green, red, and purple regions denoting veins, kidneys, arteries, and tumors, respectively

enhanced feature extraction techniques for thin structures, and more advanced post-processing methods to reduce over-segmentation.

Conclusion

In this study, RenalSegNet was introduced as a novel deep-learning framework designed to automate the segmentation of renal structures, which shows exceptional accuracy in identifying kidneys, tumors, vessels, and arteries from contrast-enhanced CT images. The proposed network's innovative architecture, including the FlexEncoder Block and MedSegPath mechanism, significantly improves feature extraction and storage of spatial data, resulting in better segmentation performance. Quantitative and qualitative analysis shows that RenalSegNet significantly outperforms current SOTA methods in various metrics, including Dice, IOU, Recall, Precision, HD, and AVD. Qualitative results further confirm the robustness and accuracy of RenalSegNet and

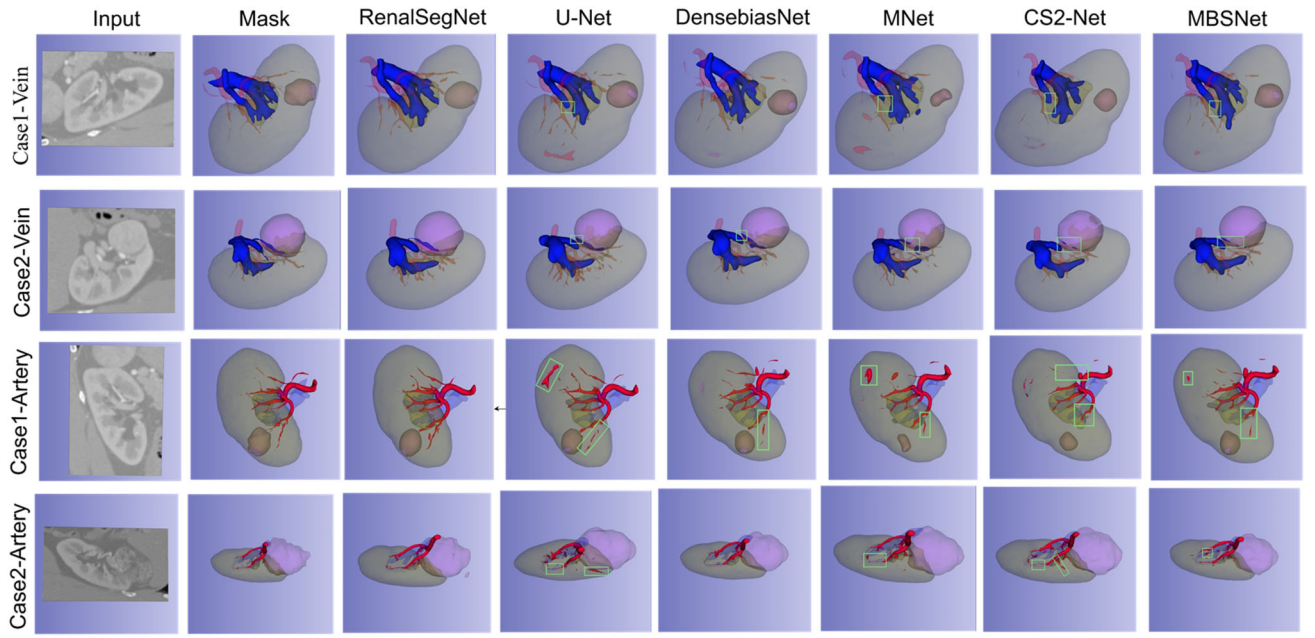

Fig. 9 The qualitative evaluation illustrates the visual superiority of the proposed model compared to five different SOTA models. Blue, green, red, and purple regions represent veins, kidneys, arteries, and tumors. Rows 1 and 2 depict the veins within the kidney structure, while Rows 3 and 4 show arteries

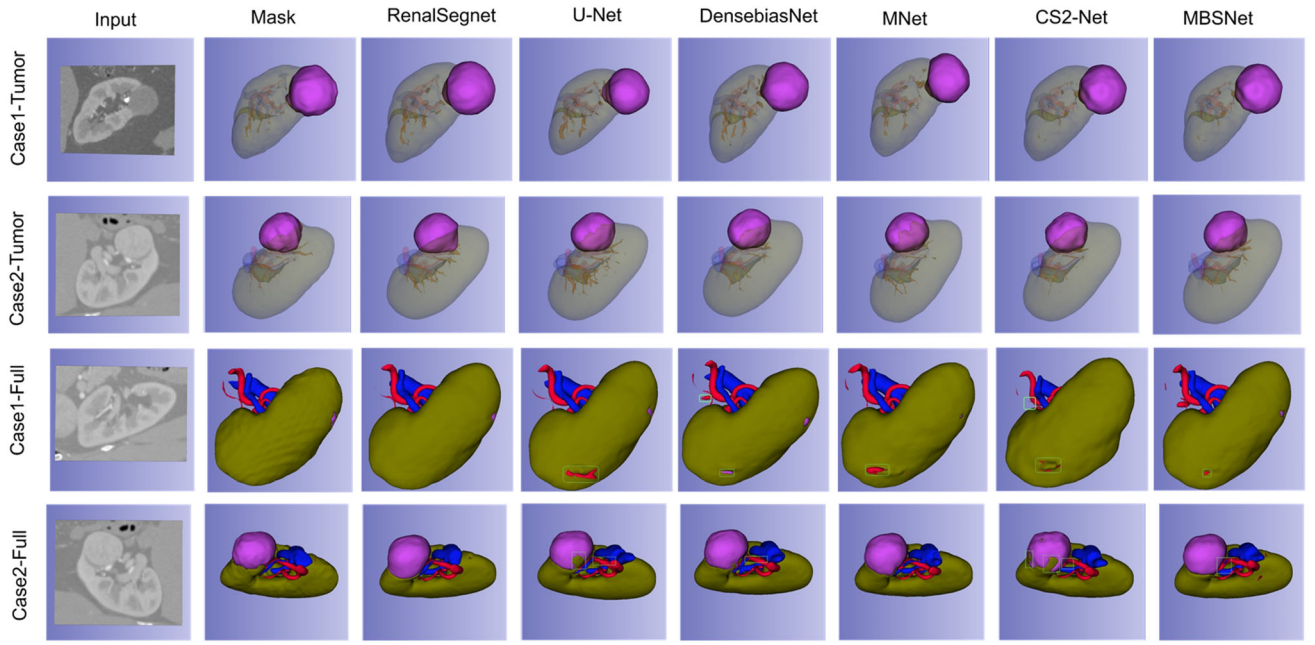

Fig. 10 The qualitative evaluation illustrates the proposed model's visual superiority compared to five state-of-the-art baseline models. Blue, green, red, and purple regions represent veins, kidneys, arteries, and tumors. Rows 1 and 2 show the tumor inside the kidney, and Rows 3 and 4 show whole kidney parts

Table 7 The ablation research illustrates the enhancements achieved to the RenalSegNet innovations on the KiPA Dataset (where ✓ denotes the inclusion of MedSegPath (MSP) or MedFuse (MF) and ✗ marks their expulsion)

MSP	MF	Average					
		Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)
✔	✔	86.25	76.75	86.69	86.48	15.78	0.79
×	✔	85.22	75.18	82.44	85.75	17.39	0.92
×	×	80.34	69.19	78.40	84.65	22.09	1.49
MSP	MF	Vein					
		Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)
✔	✔	79.99	66.65	75.31	85.39	19.07	0.90
×	✔	80.61	67.54	77.80	83.90	14.22	0.83
×	×	77.76	63.66	76.12	79.99	22.89	1.11
MSP	MF	Artery					
		Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)
✔	✔	77.45	63.54	86.19	70.38	14.64	0.53
×	✔	77.73	63.87	75.72	69.87	27.49	0.71
×	×	72.31	57.13	71.06	74.32	25.99	1.06
MSP	MF	Kidney					
		Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)
✔	✔	95.89	92.12	94.66	97.17	14.85	0.59
×	✔	95.84	92.02	95.17	96.53	18.56	0.64
×	×	95.35	91.19	96.36	94.36	25.58	0.78
MSP	MF	Tumor					
		Dice (%)	IOU (%)	Recall (%)	Precision (%)	HD (mm)	AVD (mm)
✔	✔	91.68	84.69	90.60	92.97	14.54	1.15
	✔	86.71	77.28	81.08	92.71	9.30	1.49
×	×	75.95	64.78	70.05	89.91	13.91	3.02

demonstrate its potential for clinical use in preoperative planning and postoperative evaluation. The proposed framework offers several potential advantages. It allows selective arterial compression, which improves the health of preserved renal tissue compared to complete fixation in more complex and time-consuming procedures. It also increases awareness of critical structures, such as large vessels and collecting systems near the tumor, reducing surgical complications. In addition, by superimposing tumors that cannot be seen outside the kidney, surgeons can more quickly and accurately identify tumors without the need for tactile feedback.

From a policy perspective, adopting RenalSegNet could have wide-reaching implications for healthcare systems. By improving the speed and accuracy of renal tumor segmentation, the framework could streamline surgical planning, potentially leading to reduced operating times and better patient outcomes. This improvement in efficiency could lead to better resource allocation, increasing surgical throughput, and reducing costs associated with prolonged procedures and complications. As AI-driven technologies become integral to clinical decision-making, integrating frameworks like RenalSegNet could contribute to standardizing surgical practices, thereby reducing variability and improving overall healthcare delivery.

However, despite these promising results, the model's performance is limited by the relatively small dataset and its application to a single imaging modality. Future work should address these limitations to enhance the model's robustness and generalizability. To further improve the clinical applicability and accuracy of RenalSegNet, there is a need for future work in several areas. Several areas of improvement should be explored, including integrating advanced techniques such

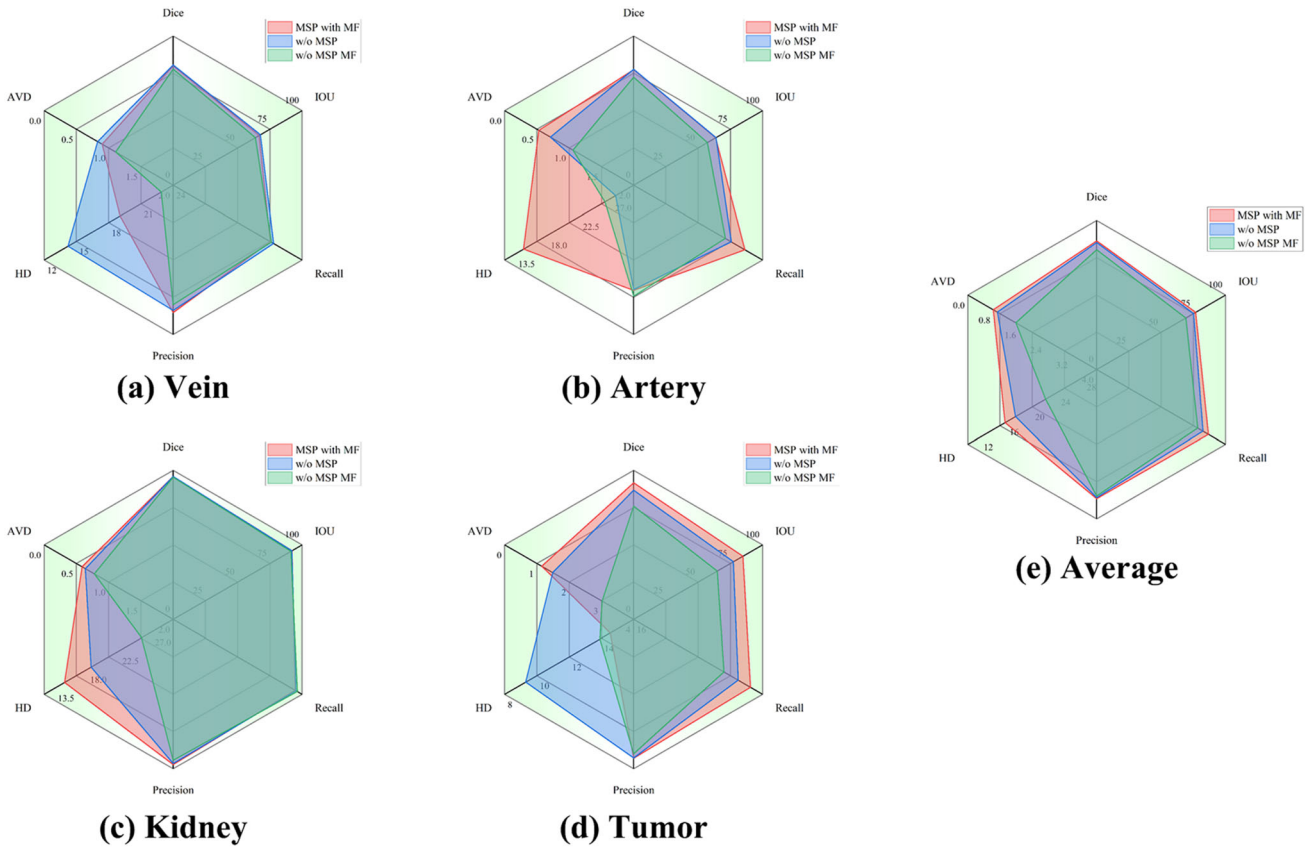

Fig. 11 RenalSegNet's performance on the KiPA Dataset is compared with (✔) and without (×) the MedSegPath (MSP) and MedFuse (MF) modules across different image segmentation metrics for **a** vein, **b** artery, **c** kidney, **d** tumor, and **e** average cases

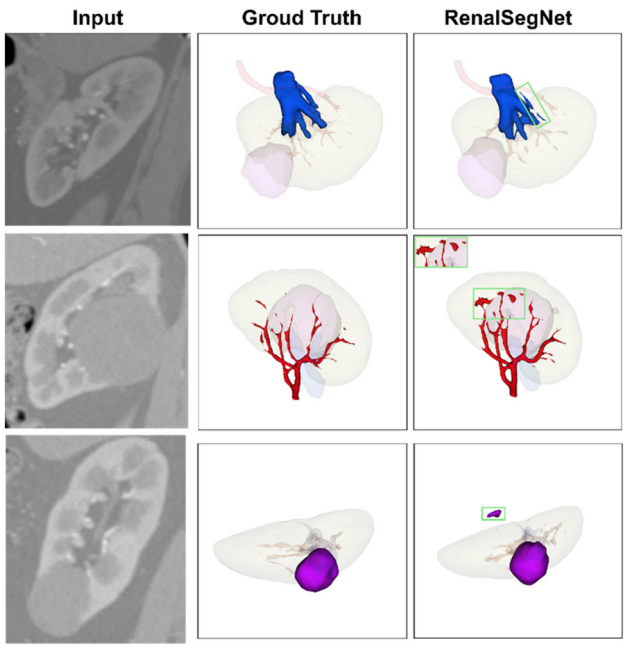

Fig. 12 Visualization of Segmentation Challenges in RenalSegNet. Difficulty in accurately segmenting low-contrast veins (Row 1), over-segmentation of small arteries (Row 2), and misclassification of tissue outside the tumor as tumor (Row 3)

as transformer-based architectures and attention mechanisms to increase the model's flexibility and accuracy. Expanding the dataset to include a wider variety of cases and imaging modalities, such as MRI and PET, will also help to improve the model's generalization and robustness. Moreover, real-time applications, especially intraoperative guidance, could be critical in complex surgeries. Finally, improving model interpretability and translating these advancements into clinical trials is essential for the broader adoption of RenalSegNet in medical practice.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40747-024-01751-2.

Acknowledgements This study was supported by the Project of the Educational Commission of Guangdong Province of China (No. 2022ZDJS113).

Author contributions Rashid Khan: Conceptualization, Methodology, Writing—review and editing. Chao Chen: Validation, Software, Formal analysis, Data curation. Asim Zaman: Visualization, Writing—review and editing, Formal analysis. Liyilei Su: Data curation, Writing—review and editing. Yan Kang: Resources, Supervision. Bingding Huang: Writing—review and editing, Supervision, Funding acquisition, Project administration.

Data availability The data presented in the figures within this paper and other study findings are available from the corresponding author (huangbingding@sztu.edu.cn) upon reasonable request. The public datasets used in this study are referenced as follows: KiPA22 [17]. The KiPA22 challenge dataset can be downloaded from the website (https://kipa22.grand-challenge.org/dataset/).

Declarations

Conflict of 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.

References

1. Williams PA, Zaidi SK, Sengupta R (2023) AACR Cancer Progress Report 2023: Advancing the frontiers of cancer science and medicine. Clin Cancer Res 29(19):3850–3851
2. Dayarathna S, Islam KT, Uribe S, Yang G, Hayat M, Chen Z (2023) Deep learning-based synthesis of MRI, CT and PET: Review and analysis. Med Image Anal 92:103046
3. Caroli A, Remuzzi A, Lerman LO (2021) Basic principles and new advances in kidney imaging. Kidney Int 100(5):1001–1011
4. Altini N, Prencipe B, Cascarano GD, Brunetti A, Brunetti G, Triggiani V, Carnimeo L et al (2022) Liver, kidney and spleen segmentation from CT scans and MRI with deep learning: a survey. Neurocomputing 490:30–53
5. Huang L, Ruan Su, Denœux T (2023) Application of belief functions to medical image segmentation: a review. Inform Fusion 91:737–756
6. Weng X, Song F, Tang M, Wang K, Zhang Y, Miao Y, Chan L-C, Lei P, Zuquan Hu, Yang F (2023) MDM-U-Net: A novel network for renal cancer structure segmentation. Comput Med Imaging Graph 109:102301
7. Punn NS, Agarwal S (2022) Modality specific U-Net variants for biomedical image segmentation: a survey. Artif Intell Rev 55(7):5845–5889
8. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D (2021) Image segmentation using deep learning: A survey. IEEE Trans Pattern Anal Mach Intell 44(7):3523–3542
9. Alnazer I, Bourdon P, Urruty T, Falou O, Khalil M, Shahin A, Fernandez-Maloigne C (2021) Recent advances in medical image processing for the evaluation of chronic kidney disease. Med Image Anal 69:101960
10. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) Unet++: a nested u-net architecture for medical image segmentation. In: 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, pp. 3–11. Springer International Publishing, 2018
11. Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK (2019) Recurrent residual U-Net for medical image segmentation. J Med imaging 6(1):014006–014006
12. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, Mori K et al (2018) Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999
13. Feng Di, Haase-Schütz C, Rosenbaum L, Hertlein H, Glaeser C, Timm F, Wiesbeck W, Dietmayer K (2020) Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. IEEE Trans Intell Transp Syst 22(3):1341–1360
14. Soydaner D (2022) Attention mechanism in neural networks: where it comes and where it goes. Neural Comput Appl 34(16):13371–13385
15. Wu H, Zhang J, Huang K, Liang K, Yu Y (2019) Fastfcn: rethinking dilated convolution in the backbone for semantic segmentation. arXiv preprint arXiv:1903.11816
16. Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, Nan Du, Fan W, Xie X (2019) AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. Med Phys 46(2):576–589
17. https://kipa22.grand-challenge.org/dataset/. Accessed 15 Mar 2024
18. Chen X, Wang X, Zhang Ke, Fung K-M, Thai TC, Moore K, Mannel RS, Liu H, Zheng B, Qiu Y (2022) Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal 79:102444
19. Khan R, Xiao C, Liu Y, Tian J, Chen Z, Su L, Li D et al (2024) Transformative deep neural network approaches in kidney ultrasound segmentation: empirical validation with an annotated dataset. Interdiscipl Sci Comput Life Sci 16:439–454
20. Zaman A, Hassan H, Zeng X, Khan R, Lu J, Yang H, Miao X et al (2024) Adaptive Feature Medical Segmentation Network: an adaptable deep learning paradigm for high-performance 3D brain lesion segmentation in medical imaging. Front Neurosci 18:1363930
21. Khan R, Su L, Zaman A, Hassan H, Kang Y, Huang B (2024) Customized m-RCNN and hybrid deep classifier for liver cancer segmentation and classification. Heliyon 10:e30528
22. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18, pp. 234–241. Springer International Publishing, 2015
23. Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, pp 2881–2890
24. Liu J-J, Hou Q, Cheng M-M, Feng J, Jiang J (2019) A simple pooling-based design for real-time salient object detection. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 3917–3926
25. Chen L-C, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587
26. Li He, Iwamoto Y, Han X, Lin L, Furukawa A, Kanasaki S, Chen Y-W (2023) 3D Multiple-Contextual ROI-Attention Network for Efficient and Accurate Volumetric Medical Image Segmentation. IEICE Trans Inf Syst 106(5):1027–1037
27. Peng C, Myronenko A, Hatamizadeh A, Nath V, Siddiquee MMR, He Y, Xu D, Chellappa R, Yang D (2022) Hypersegnas: bridging

one-shot neural architecture search with 3d medical image segmentation using hypernet. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp. 20741–20751. 2022
28. Huang Z, Wang Z, Yang Z, Gu L (2022) AdwU-Net: adaptive depth and width U-Net for medical image segmentation by differentiable neural architecture search. In: International Conference on Medical Imaging with Deep Learning, pp. 576–589. PMLR, 2022
29. Zhang Y, Li X, Chen H, Yuille AL, Liu Y, Zhou Z (2023) Continual learning for abdominal multi-organ and tumor segmentation. In: International Conference on medical image computing and computer-assisted intervention, pp. 35–45. Cham: Springer Nature Switzerland, 2023
30. Shao P, Tang L, Li P, Xu Y, Qin C, Cao Q, Ju X et al (2012) Precise segmental renal artery clamping under the guidance of dual-source computed tomography angiography during laparoscopic partial nephrectomy. Eur Urol 62(6):1001–1008
31. Wang H, Huang Z, Ye J, Tu C, Yang Y, Du S, Deng Z, Ma C, Niu J, He J (2022) An evaluation of U-Net in Renal Structure Segmentation." arXiv preprint arXiv:2209.02247
32. Peng Y, Xiqing Hu, Hao X, Liu P, Deng Y, Li Z (2024) Spider-Net: High-resolution multi-scale attention network with full-attention decoder for tumor segmentation in kidney, liver and pancreas. Biomed Signal Process Control 93:106163
33. Pan S, Chang C-W, Wang T, Wynne J, Mingzhe Hu, Lei Y, Liu T, Patel P, Roper J, Yang X (2023) Abdomen CT multi-organ segmentation using token-based MLP-Mixer. Med Phys 50(5):3027–3038
34. Isensee F, Maier-Hein KH (2019) An attempt at beating the 3D U-Net. arXiv preprint arXiv:1908.02182
35. Meyer A, Chlebus G, Rak M, Schindele D, Schostak M, van Ginneken B, Schenk A et al (2021) Anisotropic 3D multi-stream CNN for accurate prostate segmentation from multi-planar MRI. Comput Methods Programs Biomed 200:105821
36. Dong Z, He Y, Qi X, Chen Y, Shu H, Coatrieux J-L, Yang G, Li S (2022) MNet: rethinking 2D/3D networks for anisotropic medical image segmentation. In: Thirty-First International Joint Conference on Artificial Intelligence {IJCAI-22}, pp. 870–876. International Joint Conferences on Artificial Intelligence Organization, 2022
37. Feng S, Zhao H, Shi F, Cheng X, Wang M, Ma Y, Xiang D, Zhu W, Chen X (2020) CPFNet: Context pyramid fusion network for medical image segmentation. IEEE Trans Med Imaging 39(10):3008–3018
38. He Y, Yang G, Yang J, Ge R, Kong Y, Zhu X, Zhang S, Shao P, Shu H, Dillenseger JL, Coatrieux JL (2021) Meta grayscale adaptive network for 3D integrated renal structures segmentation. Med Image Anal 71:102055
39. Mou L, Zhao Y, Fu H, Liu Y, Cheng J, Zheng Y, Su P et al (2021) CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging. Med Image Anal 67:101874
40. Jin S, Sheng Yu, Peng J, Wang H, Zhao Y (2023) A novel medical image segmentation approach by using multi-branch segmentation network based on local and global information synchronous learning. Sci Rep 13(1):6762
41. Milletari F, Navab N, Ahmadi S-A (2016) V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fouRth International Conference on 3D vision (3DV), pp 565–571. Ieee
42. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18(2):203–211
43. Peng C, Myronenko A, Hatamizadeh A, Nath V, Siddiquee MMR, He Y, Xu D, Chellappa R, Yang D (2022) Hypersegnas: bridging one-shot neural architecture search with 3d medical image segmentation using hypernet. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp 20741–20751
44. Tong T, Li G, Liu X, Gao Q (2017) Image super-resolution using dense skip connections. In: Proceedings of the IEEE International Conference on computer vision, pp 4799–4807
45. He Y, Yang G, Yang J, Chen Y, Kong Y, Wu J, Tang L et al (2020) Dense biased networks with deep priori anatomy and hard region adaptation: semi-supervised learning for fine renal artery segmentation. Med Image Anal 63:101722
46. Abbasi SF, Ahmad J, Tahir A, Awais M, Chen C, Irfan M, Siddiqa HA et al (2020) Eeg-based neonatal sleep-wake classification using multilayer perceptron neural network. IEEE Access 8:183025–183034
47. Abbasi SF, Bilal M, Mukherjee T, Churm J, Pournik O, Epiphaniou G, Arvanitis TN (2024) Deep learning-based synthetic skin lesion image classification. In: Mantas J et al (eds) Digital health and informatics innovations for sustainable health care systems. IOS Press, pp 1145–50
48. Mukherjee T, Gour S, Abbasi SF, Pournik O, Arvanitis TN (2024) Development of a Cnn for adult brain tumour characterisation: implications and future directions for transfer learning. Stud Health Technol Inform 316:1674–1678