Interdisciplinary Sciences: Computational Life Sciences
https://doi.org/10.1007/s12539-024-00620-3

ORIGINAL RESEARCH ARTICLE

Transformative Deep Neural Network Approaches in Kidney Ultrasound Segmentation: Empirical Validation with an Annotated Dataset

Rashid Khan[1,2,3] · Chuda Xiao[1,4] · Yang Liu[5] · Jinyu Tian[4] · Zhuo Chen[4] · Liyilei Su[1,2,3] · Dan Li[4] · Haseeb Hassan[1] · Haoyu Li[1] · Weiguo Xie[4] · Wen Zhong[5] · Bingding Huang[1]

Received: 4 July 2023 / Revised: 6 January 2024 / Accepted: 5 February 2024

Abstract

Kidney ultrasound (US) images are primarily employed for diagnosing different renal diseases. Among them, one is renal localization and detection, which can be carried out by segmenting the kidney US images. However, kidney segmentation from US images is challenging due to low contrast, speckle noise, fluid, variations in kidney shape, and modality artifacts. Moreover, well-annotated US datasets for renal segmentation and detection are scarce. This study aims to build a novel, well-annotated dataset containing 44,880 US images. In addition, we propose a novel training scheme that utilizes the encoder and decoder parts of a state-of-the-art segmentation algorithm. In the pre-processing step, pixel intensity normalization improves contrast and facilitates model convergence. The modified encoder–decoder architecture improves pyramid-shaped hole pooling, cascaded multiple-hole convolutions, and batch normalization. The pre-processing step gradually reconstructs spatial information, including the capture of complete object boundaries, and the post-processing module with a concave curvature reduces the false positive rate of the results. We present benchmark findings to validate the quality of the proposed training scheme and dataset. We applied six evaluation metrics and several baseline segmentation approaches to our novel kidney US dataset. Among the evaluated models, DeepLabv3+ performed well and achieved the highest dice, Hausdorff distance 95, accuracy, specificity, average symmetric surface distance, and recall scores of 89.76%, 9.91, 98.14%, 98.83%, 3.03, and 90.68%, respectively. The proposed training strategy aids state-of-the-art segmentation models, resulting in better-segmented predictions. Furthermore, the large, well-annotated kidney US public dataset will serve as a valuable baseline source for future medical image analysis research.

Graphic Abstract

The graphic abstract for this research study visually encapsulates the key contributions and innovations:
Dataset creation

- Developed WD-KUS dataset (44,880 US images).
- Aims to standardize US segmentation benchmarks and simplify US interpretation efforts.

Automatic kidney segmentation framework

- Demonstrated a practical framework for segmenting whole kidneys from low-quality US images.
- Integrated various encoder–decoder models and a unique training strategy.
- Addressed challenges like shadows, internal fluid, and blurring.

Training and post-processing strategies

- Introduced effective training strategies (pre-processing, networks, learning rate).

Rashid Khan, Chuda Xiao and Yang Liu contributed equally.

Extended author information available on the last page of the article

Published online: 27 February 2024

- Utilized post-processing techniques, including concave curvature assessment.
- Improved segmentation accuracy and generalizability.

Auxiliary function for abnormal segmentation

- Proposed an auxiliary function to distinguish normal and abnormal kidney segmentation.
- Based on the concept that normal segmentation has few concave corners, while abnormal segmentation has many.

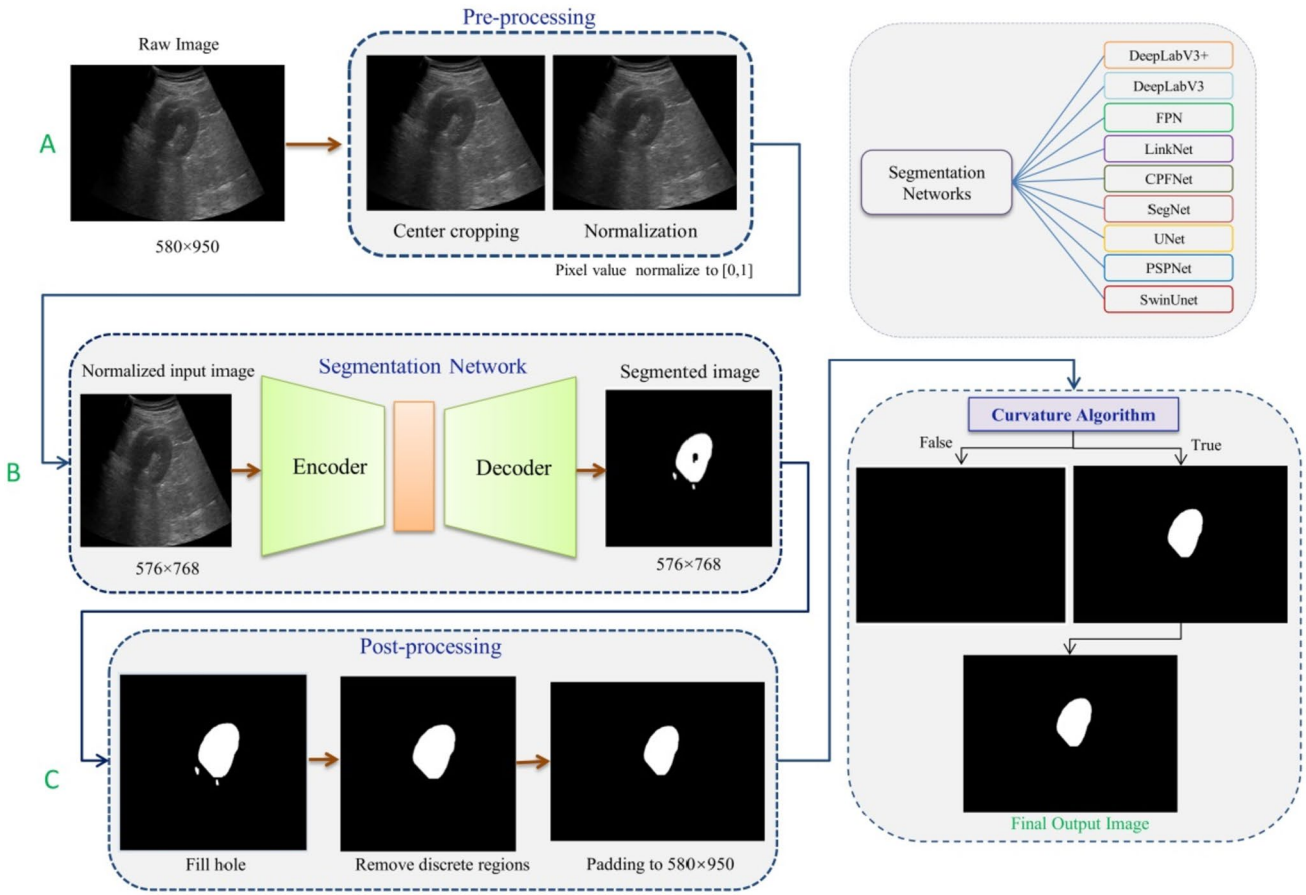
Quantitative and qualitative enhancement

- Enhanced segmentation results quantitatively (six metrics).
- Showcased qualitative improvement over baseline methods.

Validation with state-of-the-art networks

- Validated the approach using modified state-of-the-art segmentation neural networks.
- Demonstrated the effectiveness of the WD-KUS dataset in validation.

The research significantly contributes to the field by providing a comprehensive dataset, an advanced segmentation framework, and innovative strategies for training and post-processing, resulting in improved kidney segmentation accuracy and applicability.



Keywords Medical image segmentation · Kidney ultrasound · Deep learning · DeepLabv3+

1 Introduction

Ultrasound (US) is an essential component of medical imaging and one of the most widely used clinical diagnostic procedures. It is known as a robust and universal technique [1], due to its relative safety, low cost, non-invasive nature, real-time display, operator comfort, and operator experience [2], and plays a significant role in qualitative and quantitative diagnostic procedures [3]. Several studies have targeted various stages of automated medical image processing, such as data acquisition, image analysis, reconstruction, image enhancement, visualization, and management [4]. Image analysis is among the most important of these stages, combining image segmentation, registration, and quantization [5]. Medical image segmentation (MIS) is essential for effective diagnosis and clinical decision-making [6]. MIS can be described as a strategy to accurately identify the anatomy of human organs in accordance with the concerned field [7]. Since these imaging modalities are capable of detecting minor to major lesions, clinical experts recommend them for initial diagnosis.

As these US scans depict organ anatomy, they also provide insight into underlying diseases and defects [8–10]. For example, the kidney is a vital organ in the human body with an irreplaceable role. However, defective kidneys may cause severe clinical concerns and pose a high risk to public health [7]. Therefore, computer-aided diagnosis (CAD) systems are required to examine defective kidneys or renal diseases. Most CAD systems rely on image segmentation and partitioning [11]. The main objectives of kidney segmentation in medical practice are (1) to construct renal parameters, such as size and location, (2) to examine kidney anatomy and role, (3) to identify kidney abnormalities, (4) to support essential selections in the design of medical interventions, and (5) to provide post-operative support after surgical interventions [12]. However, US scans have limitations, and accurate kidney segmentation based on US scans is challenging due to low contrast, speckle noise, fluid, and variations in kidney shape [13, 14]; see Fig. 1.

Kidney segmentation has been categorized as manual, semi-automatic, and fully automatic based on current research findings. Manual kidney US (KUS) segmentation techniques are labor-intensive and time-consuming, with variations across operators to enhance their segmentation accuracy and performance. Additionally, most semi-automatic approaches address the kidney segmentation problem as a boundary detection task and focus mainly on manually initialized positions [15]. Likewise, blurred boundaries and uneven intensity distribution significantly impact the efficiency of semi-automatic segmentation methods [16]. Thus, it is invaluable to automatically extract kidneys separately from US images [17]. The application of deep learning in medical image analysis extends beyond ultrasound segmentation. It has been pivotal in fields such as radiography, pathology, and ophthalmology, aiding in anomaly detection, disease classification, and treatment outcome prediction [18, 19]. Deep learning's ability to process and discern complex patterns in medical images has replaced more labor-intensive, error-prone methods [20, 21].

Advancements in automatic image segmentation have been notably driven by the use of convolutional neural networks (CNNs) [22–24], which exhibit exceptional capabilities in learning complex, non-linear patterns in both natural and medical imaging contexts [25–27]. The U-Net architecture, with its encoder–decoder framework, has become particularly prominent in medical image segmentation (MIS) due to its effectiveness and adaptability [28]. An innovative adaptation of this approach, known as adaptive attention U-Net, has been proposed for segmenting breast lesions at various scales, showcasing the versatility of CNNs in US image segmentation [29]. Furthering this progression, an ensemble CNN-based framework has been introduced for computer-aided diagnosis (CAD) systems in breast ultrasound imaging, highlighting the continued innovation and potential of automatic segmentation methods in diverse medical imaging applications [30].

Similarly, recent advancements in novel frameworks have focused on leveraging convolutional neural networks (CNNs) for automatic segmentation of kidney ultrasound (KUS) images [13, 15, 27]. For instance, in [15], they employed a technique called boundary distance regression along with transfer learning to segment kidneys from ultrasound images. However, the effectiveness of transfer learning in extracting renal features is often constrained by optimization techniques [31, 32]. Additionally, generating precise boundary distance maps from ultrasound images with indistinct boundaries and substantial artifacts presents a formidable challenge [13, 14]. Consequently, [33] proposed the integration of boundary detection as an ancillary mechanism to enhance the efficiency of medical image segmentation (MIS) in this context.

It is known that various factors such as low image quality, variable object shapes, uneven energy distribution, blurry boundaries, and fan-shaped shadows hinder the segmentation outcomes of KUS images [13]. Consequently, newly developed segmentation methods must address the above-mentioned factors. Thus, this research proposes a pre- and post-processing approach that could be embedded easily and applied to MIS tasks. With this objective, we incorporated various encoder–decoder models, such as DeepLabV3+ [34], DeepLabV3 [35], feature pyramid network (FPN) [36], LinkNet [37], context pyramid fusion network (CPFNet) [38], semantic pixel-wise segmentation (SegNet) [39], UNet [28], pyramid scene parsing network (PSPNet) [30], and SwinUnet [40], to validate the

proposed pre- and post-processing strategy using a novel, large dataset of annotated KUS images. The main contributions of this research study are as follows:

1. A well-annotated, bulk KUS dataset containing 44,880 US images, known as WD-KUS, is constructed and made publicly available to the research community. This novel dataset may help standardize US segmentation benchmarks and, in the long term, reduce US interpretation efforts while greatly simplifying US use.
2. A practical automatic kidney segmentation framework is demonstrated by incorporating various encoder–decoder models with a proposed training strategy to segment whole kidneys from low-quality US images captured by affordable US devices. For instance, we are interested in handling complex cases, such as shadows of kidney stones and internal fluid, and resolving blurring to some extent.
3. The proposed demonstration includes training strategies (pre-processing, networks, and learning rate) and post-processing (assessing the false positive prediction using concave curvature) that can easily be embedded and applied to MIS tasks. Our proposed training scheme significantly improves the network's segmentation accuracy and generalizability. Furthermore, an auxiliary function is proposed in post-processing that helps to solve the ill-posed problem of distinguishing between normal and abnormal kidney segmentation. This function is based on the concept that normal segmentation results tend to have few or no concave corners, while abnormal segmentation results tend to have many concave corners.
4. Our designed approach enhances the segmentation results quantitatively and qualitatively upon applying six metrics and several baseline segmentation methods. Finally, modified state-of-the-art segmentation neural networks validate our annotated WD-KUS dataset's effectiveness.

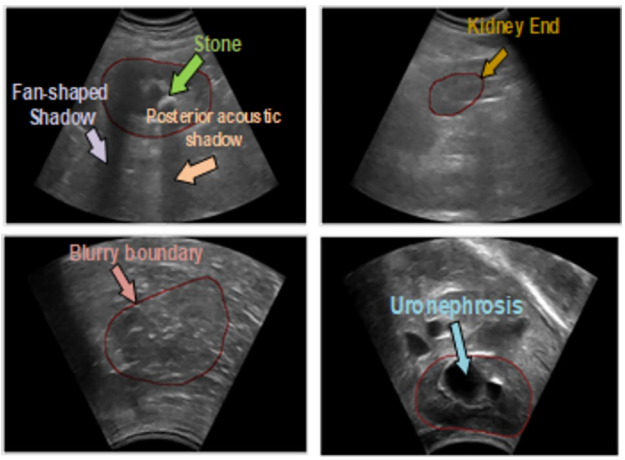


Fig. 1 Different ultrasound images are shown, with red arcs marking the borders of the kidneys. These images clearly show speckle noise, inhomogeneous intensity distribution, fan-shaped heterogeneous structures, serious cascades, and blurred boundaries

2 Materials and Methods

This section describes the dataset curation and the proposed KUS image segmentation strategy. First, the US images were acquired; then, pre-processing was performed on each volume to enhance the texture features of the image. In the second phase, a training scheme was developed using nine deep neural encoder–decoder networks, including DeepLabV3+, DeepLabV3, FPN, LinkNet, CPFNet, SegNet, UNet, PSPNet, and SwinUnet. Finally, a post-processing module was used that employed concave curvature variance (CCV) to reduce the false positive rate. Figure 2 depicts the steps involved in the development of the dataset, and Fig. 4 illustrates the steps of the proposed method.

2.1 Dataset

2.1.1 Data acquisition

A collaboration was established with the First Affiliated Hospital of Guangzhou Medical University to generate the Wuerzburg-Dynamic Kidney Ultrasound (WD-KUS) dataset from patients who had a clinical indication requiring US investigation of their kidneys. A total of 44,880 KUS images were acquired through TELEMED SmartUs EXT-1M/3M. All personal identifiable information was eliminated during data collection to ensure patients' privacy. Additionally, two different interfaces for the data were provided. At first, raw data were acquired in the MHA format via the US scan interface, followed by MP4 data via the TELEMED software.

Regarding quality, the MHA format is superior to the MP4 format. The images were collected from 148 patients. The original size of the images was set to 580×950 pixels, with the size of a pixel being 0.2745 mm. The frames were extracted from the MHA and MP4 data without resizing or cropping them and stored in the PNG imaging format. Next, patient data in the MHA format were given names such as "wdus0000-wdus-0101" and "wdus1000-wdus1049" for the MP4 format. Each patient's US has two different depths, such as 120 and 150 scans for both kidneys. The dataset included a large number of patients with kidney stones. Further details of the WD-KUS dataset are depicted in Fig. 2.

2.1.2 Data Annotation

Our data annotation procedure strictly adhered to standard annotation and medical protocols. First, a group of

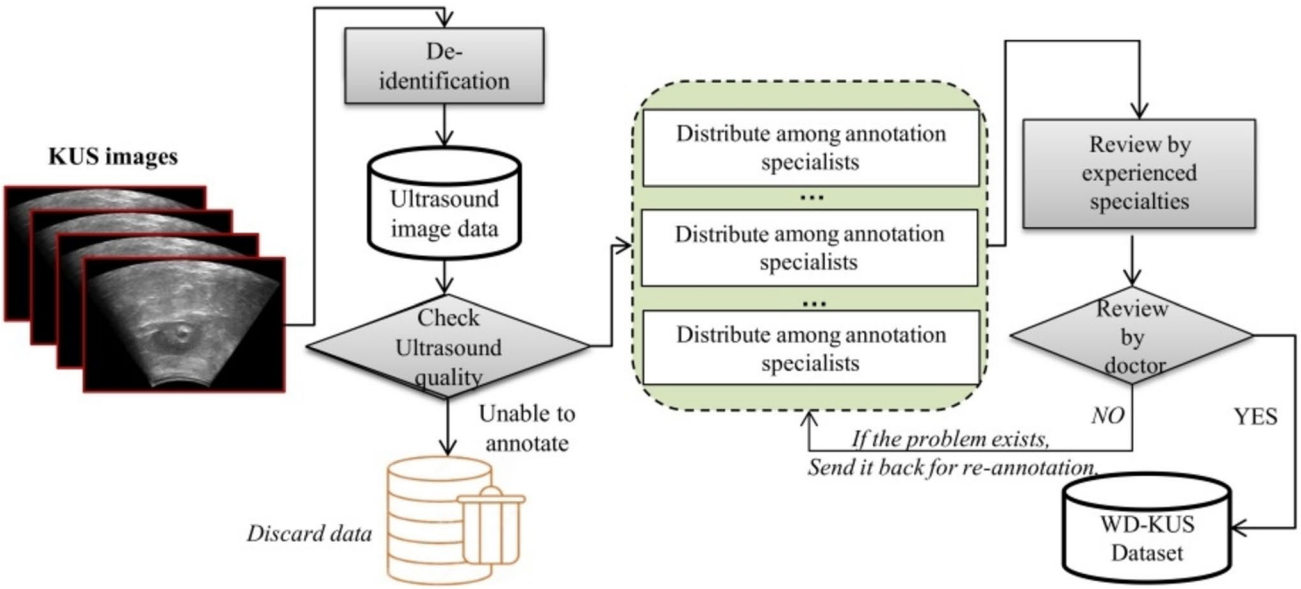


Fig. 2 The WD-KUS dataset collection's flow diagram depicts the process of gathering data

biomedical engineering and computer science students marked the data with the 3D Slicer tool. Then, the annotated data were sent to an experienced radiologist for revision. Finally, another expert urologist examined and approved the annotated data.

2.1.3 Data Augmentation

The augmentation of the WD-KUS dataset included random rotations, random scaling, random elastic deformations, gamma correction augmentation (gamma_range [0.7, 1.5]), mirroring (x- and y-axis), contrast augmentation (range 0–0.15), Gaussian noise transform (noise_variance 0.1), Gaussian blur transform (noise_variance 0.1), and simulate low-resolution transform. After applying data augmentation, a total of 44,880 KUS images were obtained. Figure 3 shows the data augmentation applied to a single image.

2.1.4 Data Splitting and Preparation

The dataset was split into training and test sets in a 4:1 ratio, with 44,880 US images from 131 patients in the training set and 33,395 images with ground truth. The test set contained 17,098 images from 32 patients and 1357 images with ground truth. The training set was further divided into training and validation sets in a 9:1 ratio. Since k-fold was used to train the network, the number of training and validation images was not fixed.

2.2 Network Architecture

The proposed KUS segmentation network was based on pre-processing, deep neural network encoder–decoder, and post-processing, as shown in Fig. 4. Our modified segmentation framework used nine different baseline models for the WD-KUS dataset. The encoder and decoder parts in these deep learning approaches for KUS image segmentation are aggregated to effectively extract and utilize both global and local image features, facilitating accurate segmentation of ultrasound images. This aggregation is key to balancing the capture of broad contextual information with the need for precise boundary and detail reconstruction in the segmented output. The adopted models are explained in detail in the following subsection.

2.2.1 Pre-processing

The original image resolution was 580×950. To improve the efficiency of network training, the images and their corresponding masks were cropped to 576×768 pixels, which reduced the number of network parameters and GPU memory usage, resulting in faster inference. In addition, pixel intensity normalization was performed, which linearly mapped the intensity values of all scans to range [0, 1] to enhance contrast and facilitate the model's convergence.

2.2.2 Encoder–Decoder Network

The encoder–decoder architecture aggregates and extracts global context information for KUS images. The decoder structure gradually reconstructs the spatial information to capture boundaries. The decoding module then recreates the boundary and location information. The proposed approach incorporates various encoder–decoder models, such as DeepLabV3 + , DeepLabV3, FPN, LinkNet, CPFNet, SegNet, UNet, PSPNet, and SwinUnet, to validate the suggested technique and our novel dataset.

2.2.3 Integration of Multiple Models in KUS Segmentation Framework

Nine deep learning algorithms were selected for the KUS dataset, and these models were trained based on the proposed pre- and post-processing approach with different settings (see Table 1). These models were classified into various types, including multiscale feature fusion models using pyramidal structures (pooling) and symmetric encoder–decoders, such as PSPNet [30], CPFNet [38], DeepLabv3+ [34], U-Net [28], SegNet [39], FPN [36], and LinkNet [37]. DeepLabv3 was used for dilated convolution to expand the perceptual field [35].

Various encoder–decoder models can be used to segment US images effectively. For instance, DeepLabV3+ is a state-of-the-art CNN that combines deep convolutional layers, atrous spatial pyramid pooling, and the encoder–decoder structure. This network can be used to perform the semantic segmentation of US images. FPN is a feature pyramid network that takes a single image of arbitrary size as input and outputs some feature maps at multiple levels in a fully convolutional fashion. Likewise, LinkNet is an encoder–decoder architecture designed for semantic segmentation tasks. Similarly, CPFNet is a context-aware segmentation model mainly designed for MIS tasks. SegNet is another type of deep convolutional encoder–decoder used for semantic segmentation problems. U-Net is another exceptional architecture designed and adopted extensively for MIS. To exploit global context information, PSPNet was proposed, which used a pyramid parsing module for global context information exploitation and outcomes in more reliable predictions. Finally, our proposed pre-and post-processing scheme was also validated with SwinUnet, a U-Net-like pure transformer for MIS.

PSPNet, CPFNet, DeepLabv3+, U-Net, SegNet, FPN, and LinkNet are all state-of-the-art models in the field of semantic segmentation, each with its own unique features. PSPNet excels in aggregating contexts at different scales, making it ideal for complex scene understanding. DeepLabv3+ combines active convolution and a fine-grained decoder for detailed segmentation at multiple scales. U-Net, known for its U-shaped architecture with jump connections, excels in medical image segmentation, especially with limited data. FPN stands out for object detection with its multiscale feature pyramid, while LinkNet's efficiency and link connectivity make it suitable for real-time applications. Each model is tailored for a specific segmentation task, ranging from detailed object profiling to efficient real-time processing.

2.2.4 Post-processing

In this step, the flood fill algorithm filled the hole and the erosion and dilation algorithm removed the discrete region [41]. Thereafter, the proposed post-processing module attempted to deploy the CCV to reduce the false positive rate of the segmented output. The mathematical formula of CCV is given below:

$$CCV = 1/N^* \sum \left(\theta_i - \mu\right)^2, \tag{1}$$

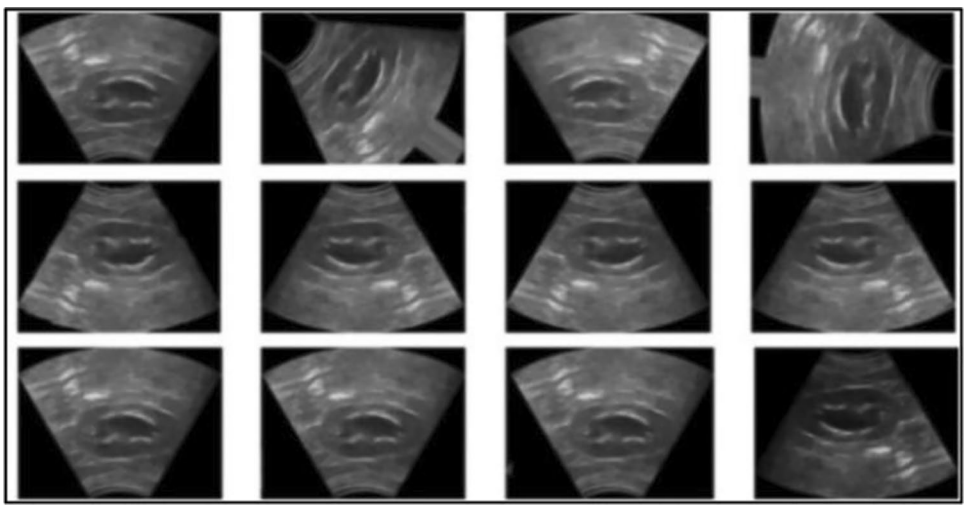


Fig. 3 Data augmentation results of a single image

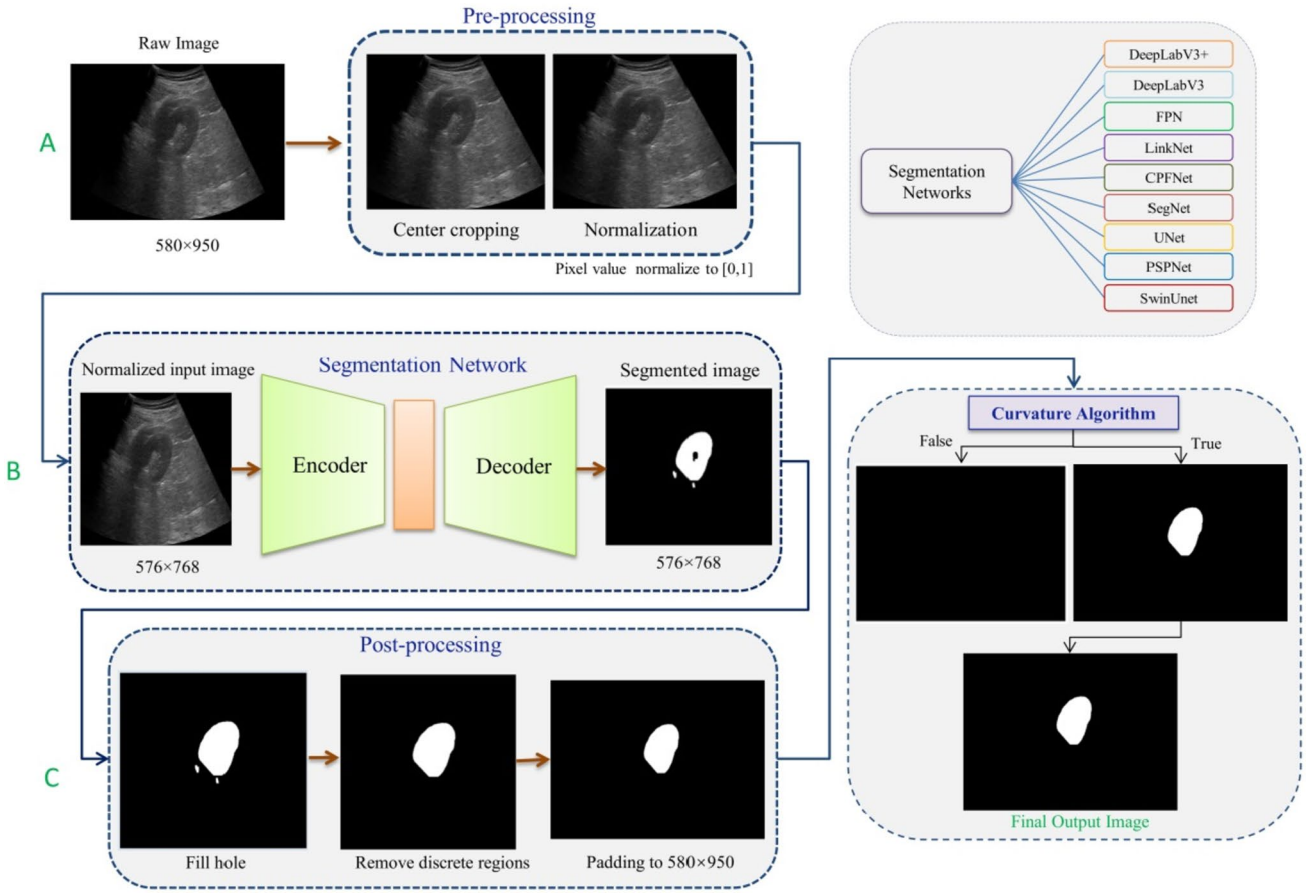


Fig. 4 The modified segmentation architecture mainly comprises: **A** Pre-processing. Initial enhancement of ultrasound images through noise reduction and contrast adjustment. **B** Segmentation networks. Utilization of nine distinct network architectures each employed sequentially for detailed feature extraction and image reconstruction. These networks include a blend of multiscale feature fusion models with pyramidal structures and symmetric encoder–decoders. **C** Post-processing phase. Refinement of segmentation outputs, focusing on minimizing false positives through the application of concave curvature analysis

where N is the number of points on the boundary, θ_i is the angle at the ith point on the boundary, and μ is the mean angle of the boundary. Equation (1) measures the degree of concavity present in the boundary, with a higher CCV value indicating a more concave shape and a lower CCV value indicating a more convex shape. The concave curvature in post-processing aims to distinguish whether kidney segmentation outcomes are well predicted. If the curvature is lower than the threshold, the predicted outcomes could be retained. This definition enables the improvement of overall model performance in practical use.

3 Experiments and Results

3.1 Implementation and Experimental Setup

In our pipeline, nine different networks were used with the annotated WD-KUS dataset to validate the framework generation and test their performance on the test set. During training, various augmentation methods, such as luminance, contrast enhancement, gamma transform, Gaussian blur, Gaussian noise, random rotation, elastic deformation, random cropping, and scaling, were randomly selected. The AdamW was selected with 100 epochs, and a batch size of 16 and weight decay of 0.001 were set to optimize the network, reduce overfitting, and improve training efficiency. Further, a combination of dice loss $\mathcal{L}_{\text{dice}}$ and focal loss $\mathcal{L}_{\text{focal}}$ with two different weights ($W1$,$W2$) was employed to train networks.

$$\mathcal{L}_{\text{total}} = W1 \times \mathcal{L}_{\text{dice}} + W2 \times \mathcal{L}_{\text{focal}}. \tag{2}$$

A weighted dice loss was chosen through quantitative evaluation of the loss function. The complete loss function of the network can be expressed as given in Eq. (3):

Table 1 The networks evaluated for kidney ultrasound segmentation

Network	No. of parameters	Backbone	Features
DeepLabV3+ [34]	22,431,442	ResNet	Atrous convolution
DeepLabV3 [35]	26,001,090	ResNet	Atrous convolution
FPN [36]	23,149,250	ResNet	Feature pyramids
LinkNet [37]	41,247,990	Efficientnet-B6	Full convolution
CPFNet [38]	43,269,891	ResNet	Context pyramid fusion
SegNet [39]	29,443,010	VGG16	Symmetric encoder–decoder
UNet [28]	2,407,442	ResNet	Symmetric encoder–decoder
PSPNet [30]	49,066,948	ResNet	Pyramid pooling
SwinUnet [40]	413,99,772	Swin	Transformer block

$$\mathcal{L}_{\text{dice}} = -\frac{2}{|P|}\sum_{k\in P}\frac{\sum_{q\in I} m_q^k n_q^k}{\sum_{q\in I} m_q^k + \sum_{q\in I} n_q^k}, \tag{3}$$

where n is a one-hot encoding of the ground truth segmentation map and m is the network's softmax output. With $q \in I$ indicating the number of pixels in the training patch/batch and $k \in P$ being the classes, both m and n have the shape $I \times P$.

Focal loss ($\mathcal{L}_{\text{focal}}$) [42] is a derivative form of cross-entropy loss (CE) that attempts to address the problem of category imbalance by assigning extra weights to challenging or easily misclassified objects, such as backgrounds with noisy textures, partial objects, or objects that are being focused on here. The focal loss is defined in Eq. (4) as follows:

$$\mathcal{L}_{\text{focal}}(p_t) = -(1-p_t)^{\gamma}\log(p_t), \tag{4}$$

where γ is the focusing parameter, and p_t is the model's estimation probability for ground truth $y \in \{\pm 1\}, p_t = \begin{cases} p & y=1 \\ 1-p & y=0 \end{cases}$. Using the focal loss can improve the stability of training when dealing with a situation where there appears to be an imbalance in the classes.

After making the inference, we applied various post-processing techniques. These included removing scatters that fell below a certain threshold, filling in any holes that were present in the segmentation, using the connect component analysis algorithm to identify the largest connected area (which we assumed corresponded to the kidney), and applying the image Concave to remove any irregularities in the shape of the image and to reduce the false positive rate [43]. The standard expression given in Eqs. (5) and (6) was used to calculate the values of S and Z, where S represents the radius and Z denotes the area of the angular surface determined by a, b, and c.

$$x(a,b,c) = \frac{1}{S}, \tag{5}$$

$$x(a,b,c) = \frac{1}{S} = \frac{4Z}{|a-b||b-c||c-a|}. \tag{6}$$

The process involves performing image segmentation to distinguish between normal and abnormal shapes based on their curvature, with the aim of automatically filtering out abnormal segmentation results by setting appropriate thresholds. PyTorch 1.9 was used as the deep learning framework, with the development environment set up on an Ubuntu 20.04.1 system featuring an AMD EPYC 7742 processor, NVIDIA A100 GPUs, and 1.8TB of RAM.

3.2 Algorithm

Algorithm 1 depicts the pseudo-code of the presented WD-KUS-based model. The network starts by considering the data set $T_r = T_1, T_2, \ldots, T_n$ to train the model. The data set $T_v = V_1, V_2, \ldots, V_n$ represents the validation set used to validate the training model, and the data set $T_s = S_1, S_2, \ldots, S_m$ was used to evaluate the model performance. The initialization weight of the network is W_0, and the best performance model weight is W_b. The threshold of the concave algorithm is Θ, and the training iterations are represented by $e = 1, 2, 3, \ldots, t$.

Algorithm 1:

	Begin:
1.	Crop all images' resolution to (576×768) in T_r;
2.	Normalize images to [0,1]
3.	**for** e ← 1 to iteration, **do**
4.	Training the network using batches in T_r
5.	Update the weights of the network $W_e \leftarrow W_{e-1}$
6.	Validate the network with weight W_e
7.	If W_e achieves better performance than W_b=W_e
8.	**end for**
9.	Using the model weights W_b inference image in T_s
10.	**for** I in T_s **do**
11.	Crop I resolution to (576×768)
12.	Normalize images to [0,1], I'
13.	Using W_b inference I' , then output O_1
14.	Fill O_1 hole in the result, then output O_2
15.	Remove O_2 discrete regions, then output O_3
16.	Padding O_3 resolution to (580×950), then output O_4
17.	Calculate the O_4 mean concave curvature $\boldsymbol{C}$
18.	**If** $C > \Theta$, then final output is O_4, then final output is None
19.	**end for**

3.3 Evaluation Metrics

To highlight the efficiency of the proposed training approach, we performed a comparative experimental analysis on the WD-KUS dataset using the six most popular assessment metrics with the baseline segmentation models [15, 44]. The six evaluation metrics included the Dice coefficient (denoted as Dice), Hausdorff distance (HD), accuracy, specificity, average symmetric surface distance (ASSD), and recall. The evaluation indicators were based on the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). A TP indicates that both the expected and actual data classes are

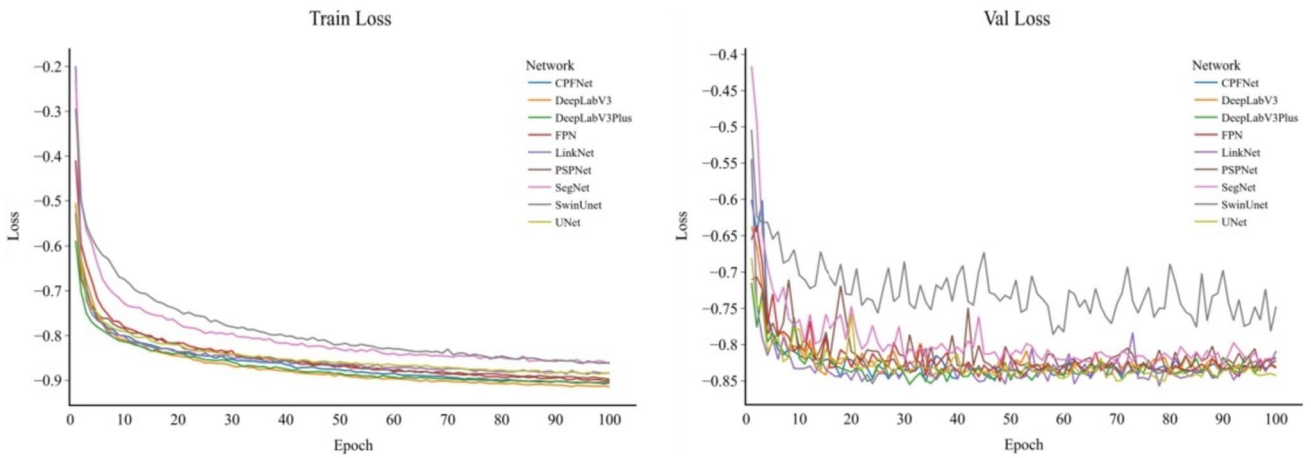


Fig. 5 Adopted networks' training and validation loss comparison

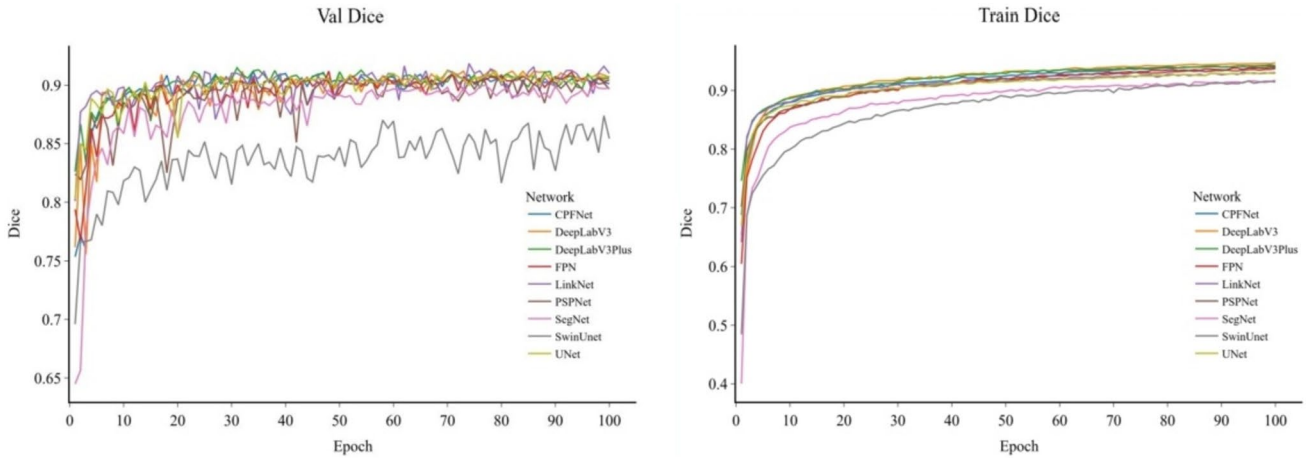


Fig. 6 Adopted networks' training and validation Dice comparison

Table 2 Quantitative evaluation of state-of-the-art segmentation methods (trained with our proposed training strategy) on the WD-KUS annotated dataset

Models	Dice (%) Mean ± std	HD95 (mm) Mean ± std	Accuracy (%) Mean ± std	Specificity (%) Mean ± std	ASSD (mm) Mean ± std	Recall (%) Mean ± std
DeepLabV3+	89.76 ± 0.62	9.91 ± 0.57	98.14 ± 0.11	98.83 ± 0.05	3.03 ± 0.19	90.68 ± 1.03
DeepLabV3	89.62 ± 0.17	9.66 ± 0.28	98.18 ± 0.04	98.91 ± 0.09	2.94 ± 0.08	90.62 ± 1.07
LinkNet	89.35 ± 0.60	8.90 ± 0.15	98.2 ± 0.04	98.85 ± 0.01	2.87 ± 0.05	91.13 ± 0.63
FPN	89.16 ± 0.25	9.89 ± 0.42	98.07 ± 0.05	98.86 ± 0.05	3.10 ± 0.09	89.99 ± 0.48
PSPNet	89.32 ± 0.90	9.669 ± 0.05	98.12 ± 0.01	98.82 ± 0.05	3.00 ± 0.01	90.85 ± 0.90
CPFNet	88.79 ± 0.79	10.54 ± 0.84	98.02 ± 0.16	98.96 ± 0.14	3.19 ± 0.25	88.92 ± 1.88
SegNet	87.02 ± 0.75	15.96 ± 1.10	97.73 ± 0.11	98.56 ± 0.06	4.18 ± 0.20	88.65 ± 0.73
UNet	88.47 ± 0.58	10.40 ± 0.35	98.00 ± 0.07	98.78 ± 0.05	3.23 ± 0.08	89.96 ± 0.28
SwinUnet	84.22 ± 0.34	13.90 ± 0.28	97.20 ± 0.05	98.44 ± 0.13	4.39 ± 0.03	85.08 ± 1.23

true. Conversely, a TN indicates that both the predicted and real data classes are false. An FP shows that the predicted data class is true but the actual data class is false. In contrast to FP, an FN indicates that while the class of expected data is false, the actual data class is true.

Dice coefficient Equation (7) describes Dice as the product of the intersection area between the predicted segment and the ground truth divided by the total number of pixels in the expected part and the ground truth image.

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

Accuracy As seen in Eq. (8), accuracy reflects the ratio of correctly predicted pixels to the total pixels.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{8}$$

Specificity Specificity is the degree of background that was correctly predicted and is calculated using Eq. (9).

$$\text{Specificity} = \frac{T_\text{N}}{T_\text{N} + F_\text{P}}. \tag{9}$$

Recall Recall or sensitivity is the ratio of accurately predicted foreground to all ground truth foreground and is calculated using Eq. (10).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{10}$$

ASSD The ASSD is the sum of all distances between points on the boundary of the machine-segmented region and the ground truth boundary.

$$ASSD = \frac{\sum_{a\in A} \min_{b\in B} \parallel a - b \parallel + \sum_{b\in B} \min_{a\in A} \parallel a - b \parallel}{N_A + N_B}, \tag{11}$$

where A and B represent the boundaries of segmented and reference images, and a and b denote locations on A and B accordingly. The distance between a and b is indicated by $\parallel a - b \parallel$. The numbers N_A and N_B refer to the number of positions on A and B.

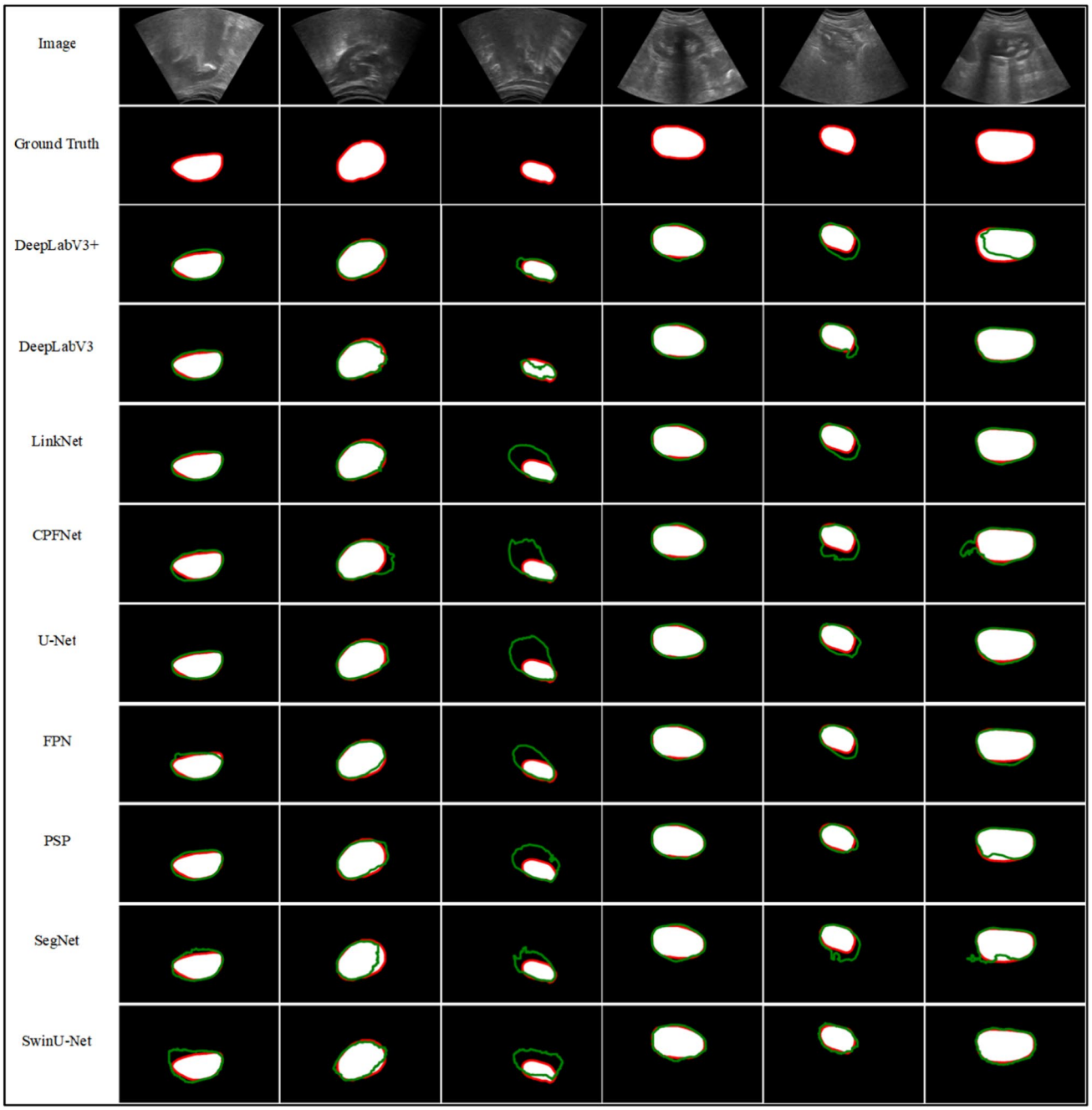


Fig. 7 KUS images' segmented maps. From top to bottom: input image, ground truth, DeeplabV3+, DeepLabV3, LinkNet, CPFNet, U-Net, FPN, PSPNet, SegNet, and SwinUNet. The areas highlighted with green and red contours indicate the ground truth and the prediction outcomes, respectively

Hausdorff distance (HD95) HD95 is comparable to max HD. However, it is based on the 95th percentile of the distance between X and Y boundary points. This indicator was employed to reduce the effect of a small number of outliers.

$$HD_{95}(A,B) = \max\left(d_{95}(A,B), d_{95}(B,A)\right), \tag{12}$$

$$d_{95}(A,B) = \max\left(K^{95}\begin{pmatrix}\mathrm{dis}(a,B)\\ a\in A\end{pmatrix}\right), \tag{13}$$

$$\mathrm{dis}(a,B) = \underbrace{\min}_{b\in B}||a-b||, \tag{14}$$

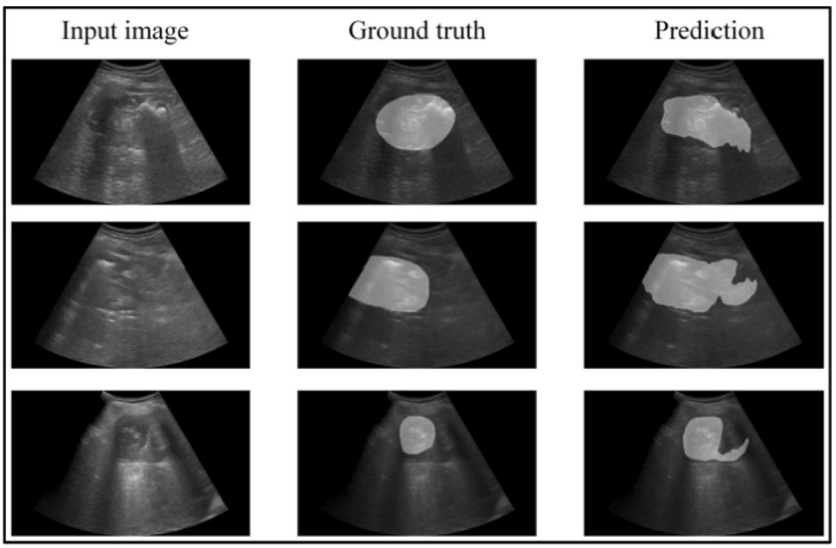


Fig. 8 Effectiveness of CCV in mitigating hydronephrosis artifacts in kidney ultrasound images. This figure visually demonstrates the impact of the curvature constraint value method in filtering out hydronephrosis effects from KUS images

Table 3 The kidney average removal percentage

Threshold	Dice (%) Mean ± std	ASSD (mm) Mean ± std	Overall removed (%)	Per kidney avg removed (%)
0.013	91.8 ± 0.139	1.10 ± 1.36	42.5	40
0.015	91.8 ± 0.136	1.13 ± 1.26	37.4	35
0.018	91.6 ± 0.135	1.18 ± 1.34	32.2	30
0.023	91.4 ± 0.136	1.22 ± 1.39	27	25
0.029	**91.2** ± 0.136	**1.28** ± 1.55	**21.6**	**20**
0.038	90.9 ± 0.138	1.35 ± 1.70	16.5	15
0.054	90.6 ± 0.143	1.42 ± 1.89	11.3	10

The bold values represent the selected threshold value and its corresponding performance metrics

where A and B each provide a different point set. K^{95} represents the 95th percentile of distances. ASSD measures the average mutual distance between the two surfaces' edges, whereas HD is the maximal symmetric surface distance. The ASSD and HD were determined using Eq. (11) and Eqs. (12–14), respectively. Figure 5 shows the training and validation loss traces, and Fig. 6 shows the training and validation Dice for the KUS network trained with the proposed pre-processing and post-processing training strategy.

3.4 Ablation Analysis and Comparisons

This study conducted comparative experiments with widely used MIS networks, namely DeepLabV3+, DeepLabV3, FPN, LinkNet, CPFNet, SegNet, UNet, PSPNet, and SwinUnet. Table 2 shows the quantitative analysis regarding six evaluation metrics, where our proposed DeeplabV3+ encoder–decoder strategy achieves the best results. The ablation experiments were mainly performed on our annotated WD-KUS dataset. The effectiveness of the proposed pre-processing and post-processing training strategy was demonstrated by nine different state-of-the-art segmentation approaches.

Currently, FPN, DeepLabV3, DeeplabV3+, SegNet, and PSPNet are widely used in comparative image segmentation experiments. Therefore, we mainly used these methods to compare our experimental results. In addition, other state-of-the-art segmentation models were considered for training and evaluation purposes, such as CPFNet, UNet, and CPFNet. The final predicted results of DeepLabV3, Deeplabv3+, LinkNet, PSPNet, and CPFNet were notably better with the proposed training scheme on our annotated WD-KUS dataset. We also noted that the incorporated pre-training and post-training steps with adopted segmentation frameworks generated better-segmented outcomes even from the US images with unclear boundaries. Figure 7 shows the segmented maps of the sample images obtained from the test dataset using the proposed training method.

Moreover, we utilized the CCV during post-processing to identify false positive predictions. The purpose of CCV in post-processing was to distinguish whether kidney segmentation results are well predicted. However, if the curvature was lower than the threshold, then we retained the previous segmented results. This technique allows us to improve the overall performance of the models in practical use [45]. Figure 8 shows the effectiveness of the CCV method by using it to remove the effects of hydronephrosis in US images of the kidney.

Table 3 shows the post-processing inference results to make the segmentation shape more regular. After choosing a

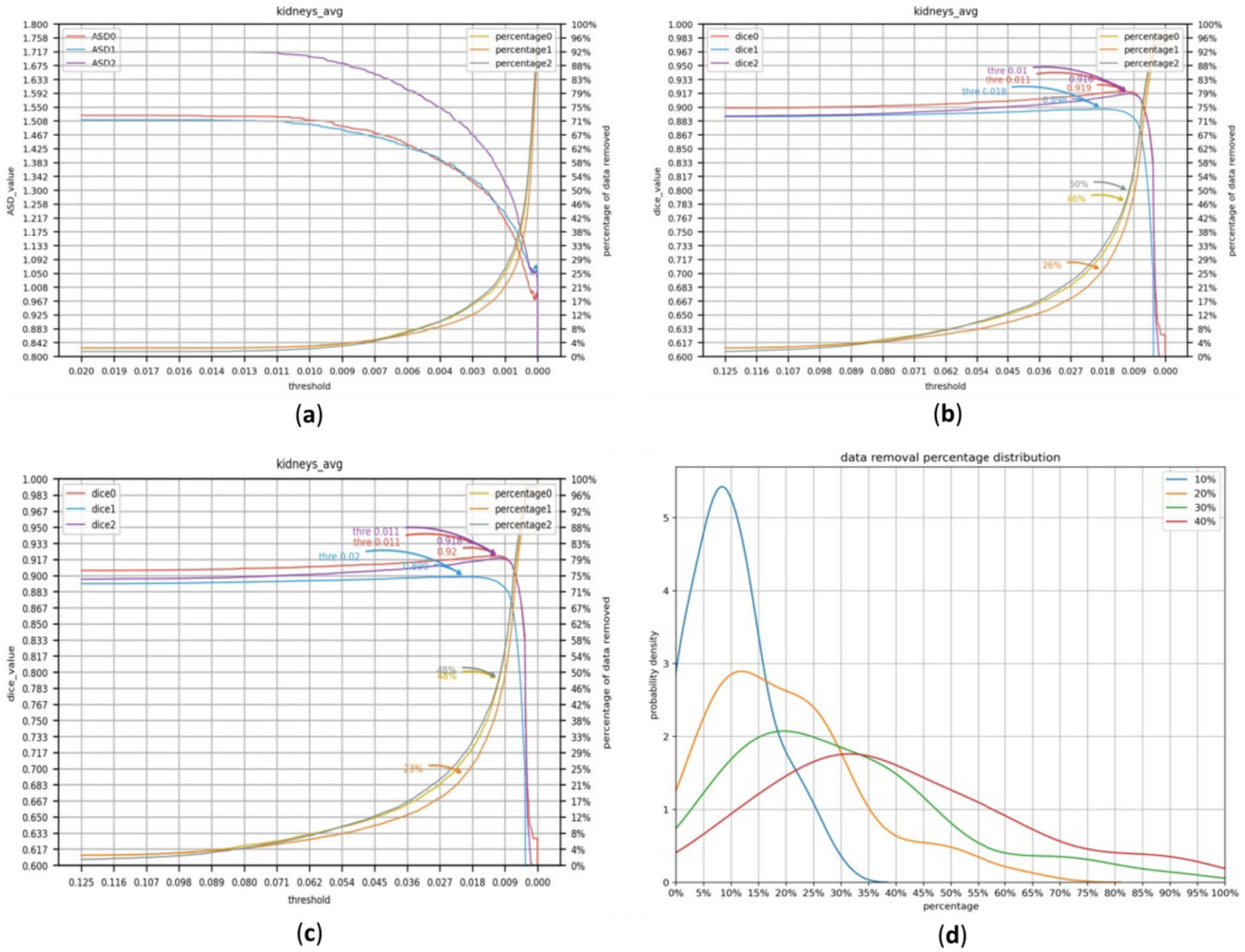


Fig. 9 Analysis of the thresholds: **a** the ASSD plot threshold is decreasing (indicate that the average removal rate of each intrarenal image increases as the curvature threshold decreases); **b** the Dice threshold plot is exclusive to the model's prediction; **c** the Dice threshold value is exclusive to predictions with ground truth; **d** per kidney data removal percentage distribution. The arrows with percentages correspond to their highest score

threshold value of 0.029 in the depression algorithm and not calculating the performance of the image if the depression value was more significant than the threshold, we observed that the average renal removal percentage became increasingly sensitive to threshold changes.

The distribution of each kidney prediction removal percentage in the test set is shown. Hydronephrosis is characterized by high kidney removal rates, as shown in Table 3. For 10%, 20%, 30%, 40%, and 50%, this indicates that the average removal rate of the kidneys is less than 10%. The same is true for 20%, 30%, 40%, and 50%. These values considered different levels of kidney prediction removal during post-processing, representing the threshold percentages below which the average removal rate of the kidneys is observed.

As shown in Fig. 9, the threshold decreased, and the ASSD went lower, which is preferable. The Dice score increased with a decreasing threshold. The arrows in Fig. 9c indicate the highest Dice score. The highest Dice value for Dice 0 models was at a threshold of 0.011. The plot in Fig. 9c shows that US images without ground truth are either challenging to annotate or there is no kidney. The differences between plots 9b and 9c should be minimal if the curvature method is successful, and Fig. 9d shows high removal rates in the kidneys. The proposed technique was developed in such a manner that the lower the threshold, the more segmentation outcomes are eliminated.

The experimental results showed that DeepLabv3+, with the proposed pre- and post-processing methods, achieved

the highest segmentation accuracy for segmenting the kidney based on the novel WD-KUS imaging dataset. Specifically, DeepLabv3+ scored 89.76%, 9.91, 3.03, 98.14%, and 90.68% for Dice, HD95, ASSD accuracy, and recall, respectively. Based on the evaluation of Table 2 and Fig. 7, it can be concluded that the modified segmented framework has a significant competitive advantage and significantly decreases the false and missed detection rates compared to other segmentation approaches.

4 Conclusion

Accurate and automated segmentation of kidneys in KUS images is essential for reliable clinical diagnosis and treatment. Recently, much attention has been paid to automatic kidney segmentation due to underlying challenges such as the poor quality of images, varying sizes and shapes of kidneys, potential presence of kidney stones, fluids, and heterogeneous structures. Moreover, KUS segmentation data are rarely public and mostly restricted to private datasets. Thus, this study proposed a suitable training strategy and the curation of a novel, well-annotated WD-KUS dataset containing 44,880 US images. Our modified architecture combines pre-processing, deep neural encoders-decoders, and post-processing (concave curvature) for accurate and robust kidney segmentation to address the challenges in KUS image segmentation. We incorporated our proposed strategy into nine baseline segmentation encoder–decoder models and validated the modified segmentation architectures using our in-house annotated dataset. Using the proposed training technique, all the segmentation algorithms performed well regarding Dice, specificity, sensitivity, and accuracy. The findings demonstrate the superiority of the DeepLabv3+ encoder–decoder, particularly when augmented with the proposed pre- and post-processing methods, in the context of kidney segmentation using the WD-KUS dataset. Notably, DeepLabv3+ achieved impressive performance metrics, with a Dice score of 89.76%, Hausdorff distance (HD95) of 9.91, average symmetric surface distance (ASSD) of 3.03, accuracy of 98.14%, and a recall of 90.68%. The experimental results showed that our modified segmentation networks achieved very competitive segmented outcomes regarding Dice, specificity, sensitivity, and accuracy. However, we need to overcome the challenges posed by factors such as kidney stones, which cause a posterior hypoechoic shadow, and shadows caused by ribs, leading to segmentation uncertainty and substandard results. This research can potentially influence the development of automated diagnostic tools in other areas of medical imaging, leveraging deep learning for broader clinical impact.

Acknowledgements This study was supported by the Project of the Educational Commission of Guangdong Province of China (No. 2022ZDJS113). We would also like to thank the School-Enterprise Graduate Student Cooperation Fund of Shenzhen Technology University and the School-Enterprise Cooperation Fund provided by Wuerzburg Dynamics, Inc., to the Weiding Joint Laboratory of Medical Artificial Intelligence, Shenzhen Technology University.

Data Availability The data presented in the figures within this paper and other finding of this study are available from the corresponding authors upon reasonable request.

Code Availability The custom codes used to produce the results presented in this paper are available from the corresponding authors upon reasonable request.

Declarations

Conflict of Interest The authors state that they have no known competing financial interests or close personal ties that could have influenced the research work presented in this paper.

References

1. Rahman T, Khandakar A, Qiblawey Y et al (2021) Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Comput Biol Med 132:104319. https://doi.org/10.1016/j.compbiomed.2021.104319
2. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25(1):44–56. https://doi.org/10.1038/s41591-018-0300-7
3. Li P, Zhao H, Liu P et al (2020) Automated measurement network for accurate segmentation and parameter modification in fetal head ultrasound images. Med Biol Eng Comput 58:2879–2892. https://doi.org/10.1007/s11517-020-02242-5
4. Kim T-Y, Son J, Kim K-G (2011) The recent progress in quantitative medical image analysis for computer aided diagnosis systems. Healthc Inform Res 17(3):143–149. https://doi.org/10.4258/hir.2011.17.3.143
5. Huang Y, Yang X, Liu L et al (2023) Segment anything model for medical images? Med Image Anal. https://doi.org/10.1016/j.media.2023.103061
6. Shi F, Wang J, Shi J et al (2020) Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE Rev Biomed Eng 14:4–15. https://doi.org/10.1109/RBME.2020.2987975
7. Levey AS, Coresh J (2012) Chronic kidney disease. Lancet 379(9811):165–180. https://doi.org/10.1016/S0140-6736(11)60178-5
8. Yin S, Zhang Z, Li H et al (2019) Fully-automatic segmentation of kidneys in clinical ultrasound images using a boundary distance regression network. In: IEEE ISBI 2019, pp 1741–1744. https://doi.org/10.1109/ISBI.2019.8759170
9. Liu Y (2006) Renal fibrosis: new insights into the pathogenesis and therapeutics. Kidney Int 69(2):213–217. https://doi.org/10.1038/sj.ki.5000054
10. Wolf G, Ritz E (2005) Combination therapy with ACE inhibitors and angiotensin II receptor blockers to halt progression of chronic renal disease: pathophysiology and indications. Kidney Int 67(3):799–812. https://doi.org/10.1111/j.1523-1755.2005.00145.x
11. Nayantara PV, Kamath S, Manjunath K et al (2020) Computer-aided diagnosis of liver lesions using CT images: a systematic

review. Comput Biol Med 127:104035. https://doi.org/10.1016/j.compbiomed.2020.104035
12. Torres HR, Queiros S, Morais P et al (2018) Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: a systematic review. Comput Methods Programs Biomed 157:49–67. https://doi.org/10.1016/j.cmpb.2018.01.014
13. Chen G, Yin J, Dai Y et al (2022) A novel convolutional neural network for kidney ultrasound images segmentation. Comput Methods Programs Biomed 218:106712. https://doi.org/10.1016/j.cmpb.2022.106712
14. Torres HR, Queirós S, Morais P et al (2020) Kidney segmentation in 3-D ultrasound images using a fast phase-based approach. IEEE Trans Ultrason Ferroelectr Freq Control 68(5):1521–1531. https://doi.org/10.1109/TUFFC.2020.303933
15. Yin S, Peng Q, Li H et al (2020) Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks. Med Image Anal 60:101602. https://doi.org/10.1016/j.media.2019.101602
16. Militello C, Rundo L, Toia P et al (2019) A semi-automatic approach for epicardial adipose tissue segmentation and quantification on cardiac CT scans. Comput Biol Med 114:103424. https://doi.org/10.1016/j.compbiomed.2019.103424
17. Xie J, Jiang Y, Tsui H-t (2005) Segmentation of kidney from ultrasound images based on texture and shape priors. IEEE Trans Med Imaging 24(1):45–57. https://doi.org/10.1109/TMI.2004.837792
18. Wu H, Zhou H, Zhou B et al (2023) SCMcluster: a high-precision cell clustering algorithm integrating marker gene set with single-cell RNA sequencing data. Brief Funct Genom. https://doi.org/10.1093/bfgp/elad004
19. Zhang P, Wu Y, Zhou H et al (2022) CLNN-loop: a deep learning model to predict CTCF-mediated chromatin loops in the different cell lines and CTCF-binding sites (CBS) pair types. Bioinformatics 38(19):4497–4504. https://doi.org/10.1093/bioinformatics/btac575
20. Zhang P, Zhang H, Wu H (2022) iPro-WAEL: a comprehensive and robust framework for identifying promoters in multiple species. Nucleic Acids Res 50(18):10278–10289. https://doi.org/10.1093/nar/gkac824
21. Zhang P, Wu H (2023) ICHROM-deep: an attention-based deep learning model for identifying chromatin interactions. IEEE J Biomed Health Inform. https://doi.org/10.1109/JBHI.2023.3292299
22. Minaee S, Boykov YY, Porikli F et al (2021) Image segmentation using deep learning: a survey. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TMI.2021.3089292
23. Liu S, Qi L, Qin H et al (2018) Path aggregation network for instance segmentation. In: IEEE CVPR 2018. pp 8759–8768. https://doi.org/10.48550/arXiv.1803.01534
24. Noh H, Hong S, Han B ()2015 Learning deconvolution network for semantic segmentation. In: IEEE ICCV 2015. pp 1520–1528. https://doi.org/10.48550/arXiv.1505.04366
25. Zhang Z, Sejdić E (2019) Radiological images and machine learning: trends, perspectives, and prospects. Comput Biol Med 108:354–370. https://doi.org/10.1016/j.compbiomed.2019.02.017
26. Sarker MMK, Rashwan HA, Akram F et al (2018) SLSDeep: skin lesion segmentation based on dilated residual and pyramid pooling networks. In: MICCAI 2018. Springer, pp 21–29. https://doi.org/10.1007/978-3-030-00934-2_3
27. Hatamizadeh A, Tang Y, Nath V et al (2022) Unetr: transformers for 3D medical image segmentation. In: WACV 2022. pp 574–584. https://doi.org/10.48550/arXiv.2103.10504
28. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: MICCAI 2015. Springer, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
29. Chen G, Li L, Dai Y et al (2022) AAU-net: an adaptive attention u-net for breast lesions segmentation in ultrasound images. IEEE Trans Med Imaging. https://doi.org/10.1109/TMI.2022.3226268
30. Zhao H, Shi J, Qi X et al (2017) Pyramid scene parsing network. In: IEEE CVPR 2017. pp 2881–2890. https://doi.org/10.48550/arXiv.1612.01105
31. Chan H-P, Samala RK, Hadjiiski LM et al (2020) Deep learning in medical image analysis. Deep learning in medical image analysis: challenges and applications. Springer, Cham, pp 3–21. https://doi.org/10.1007/978-3-030-33128-3_1
32. Fiorentino MC, Villani FP, Di Cosmo M et al (2023) A review on deep-learning algorithms for fetal ultrasound-image analysis. Med Image Anal 83:102629. https://doi.org/10.1016/j.media.2022.102629
33. Arbelaez P, Maire M, Fowlkes C et al (2010) Contour detection and hierarchical image segmentation. IEEE Trans Pattern Anal Mach Intell 33(5):898–916. https://doi.org/10.1109/TPAMI.2010.161
34. Chen L-C, Zhu Y, Papandreou G et al (2018) Encoder–decoder with atrous separable convolution for semantic image segmentation. In: ECCV 2018. pp 801–818. https://doi.org/10.48550/arXiv.1802.02611
35. Chen L-C, Papandreou G, Schroff F et al (2017) Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:170605587. https://doi.org/10.48550/arXiv.1706.05587
36. Lin T-Y, Dollár P, Girshick R et al (2017) Feature pyramid networks for object detection. In: Proceedings of IEEE conference CVPR 2017. pp 2117–2125. https://doi.org/10.48550/arXiv.1612.03144
37. Chaurasia A, Culurciello E (2017) Linknet: exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE visual, pp 1–4. https://doi.org/10.1109/VCIP.2017.8305148
38. Feng S, Zhao H, Shi F et al (2020) CPFNet: Context pyramid fusion network for medical image segmentation. IEEE Trans Med Imaging 39(10):3008–3018. https://doi.org/10.1109/TMI.2020.2983721
39. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder–decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615
40. Cao H, Wang Y, Chen J et al (2022) Swin-unet: Unet-like pure transformer for medical image segmentation. In: ECCV 2022. Springer, pp 205–218. https://doi.org/10.1007/978-3-031-25066-8_9
41. Nosal E-M (2008) Flood-fill algorithms used for passive acoustic detection and tracking. In: IEEE ENVIRONMENT 2008. pp 1–5. https://doi.org/10.1109/PASSIVE.2008.4786975
42. Lin T-Y, Goyal P, Girshick R et al (2017) Focal loss for dense object detection. In: IEEE ICCV 2017. https://doi.org/10.48550/arXiv.1708.02002
43. Marciniak M, Gilbert A, Loncaric F et al (2021) Septal curvature as a robust and reproducible marker for basal septal hypertrophy. J Hypertens 39(7):1421. https://doi.org/10.1097/HJH.0000000000002813
44. Xue C, Zhu L, Fu H et al (2021) Global guidance network for breast lesion segmentation in ultrasound images. Med Image Anal 70:101989. https://doi.org/10.1016/j.media.2021.101989
45. Li S, Jin J, Daly I et al (2022) Feature selection method based on Menger curvature and LDA theory for a P300 brain–computer interface. J Neural Eng 18(6):066050. https://doi.org/10.1088/1741-2552/ac42b4

Authors and Affiliations

Rashid Khan[1,2,3] · Chuda Xiao[1,4] · Yang Liu[5] · Jinyu Tian[4] · Zhuo Chen[4] · Liyilei Su[1,2,3] · Dan Li[4] · Haseeb Hassan[1] · Haoyu Li[1] · Weiguo Xie[4] · Wen Zhong[5] · Bingding Huang[1]

✉ Wen Zhong
gzgyzhongwen@163.com

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] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University Health Science Center, Shenzhen 518060, China

[4] Wuerzburg Dynamics Inc., Shenzhen 518188, China

[5] Department of Urology, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China