Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

A novel convolutional neural network for kidney ultrasound images segmentation

Gongping Chen[a], Jingjing Yin[a], Yu Dai[a,*], Jianxun Zhang[a], Xiaotao Yin[b], Liang Cui[c]

[a] The Institute of Robotics and Automatic Information System, Tianjin Key Laboratory of Intelligent Robotics, College of Artificial Intelligence, Nankai University, Tianjin 300350, China
[b] Department of Urology, Fourth Medical Center of Chinese PLA General Hospital, Beijing 10048, China
[c] Department of Urology, Civil Aviation General Hospital, Beijing 100123, China

ARTICLE INFO

Article history:
Received 23 November 2021
Revised 27 January 2022
Accepted 22 February 2022

Keywords:
Kidney ultrasound
Automatic segmentation
Deep supervised
Deep learning

ABSTRACT

Background and Objective: Ultrasound imaging has been widely used in the screening of kidney diseases. The localization and segmentation of the kidneys in ultrasound images are helpful for the clinical diagnosis of diseases. However, it is a challenging task to segment the kidney accurately from ultrasound images due to the interference of various factors.
Methods: In this paper, a novel multi-scale and deep-supervised CNN architecture is proposed to segment the kidney. The architecture consists of an encoder, a pyramid pooling module and a decoder. In the encoder, we design a multi-scale input pyramid with parallel branches to capture features at different scales. In the decoder, a multi-output supervision module is developed. The introduction of the multi-output supervision module enables the network to learn to predict more precise segmentation results scale-by-scale. In addition, we construct a kidney ultrasound dataset, which contains of 400 images and 400 labels.
Results: To highlight effectiveness of the proposed approach, we use six quantitative indicators to compare with several state-of-the-art methods on the same kidney ultrasound dataset. The results of our method on the six indicators of accuracy, dice, jaccard, precision, recall and ASSD are 98.86%, 95.86%, 92.18%, 96.38%, 95.47% and 0.3510, respectively.
Conclusions: The analysis of evaluation indicators and segmentation results shows that our method achieves the best performance in kidney ultrasound image segmentation.

1. Introduction

The kidney is one of the most important organs for maintaining the balance of the body, and its role cannot be replaced. However, kidney disease and its complications are medical problems for human beings, which seriously threaten human health [1]. Due to the early asymptomatic, concealed nature and many predisposing factors of kidney diseases, early screening is very important for the diagnosis of kidney diseases. Segmentation is a common method in medical image analysis, which can help characterize tissues and improve diagnosis by segmenting useful targets or regions [2]. Specially, the segmentation of kidney ultrasound (KUS) images can not only evaluate the parameters, morphology and function of the kidney, but also promote the decision-making process [1]. However, due to the influence of kidney morphology, heterogeneous structure, image quality and so on, it is still a challenge to segment the kidney accurately from the ultrasound image, as shown in Fig. 1.

In the past ten years, there have been many researches on segmenting kidneys from ultrasound images, and these methods can be roughly divided into three types: manual, semiautomatic and automatic [1,2]. Manual segmentation is the most primitive method and involves an expert manually delineating the outline of the kidney from ultrasound images. However, affected by the experience of doctors and the quality of ultrasound images, manual annotation methods often cause inevitable errors [1]. In addition, manual segmentation is a very time-consuming task. In order to alleviate this challenge, a number of semiautomatic segmentation methods based on images, variable volume models, and shape models have been proposed [3–7]. Zheng et al. [3] developed a graph cuts based method to segment KUS images by integrating original image intensity information and texture feature maps extracted using Gabor filters. However, this method is susceptible to shadow and speckle noise. To alleviate the disturbance

[*] Corresponding author.
E-mail addresses: cgp110@stu.xhu.edu.cn (G. Chen), daiyu@nankai.edu.cn (Y. Dai).

https://doi.org/10.1016/j.cmpb.2022.106712
0169-2607/© 2022 Elsevier B.V. All rights reserved.

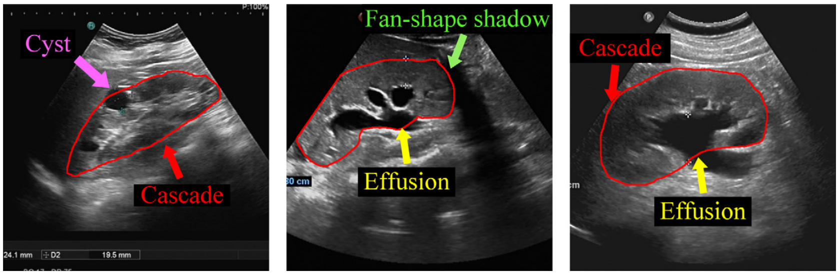

Fig. 1. Various kidney ultrasound images. The orange curve is the boundary of the kidney. Obvious speckle noise, uneven intensity distribution, fan-shaped, heterogeneous structure, serious cascade and blurred boundary can be seen from these images.

of the segmentation results by noise, the segmentation method based on variable volume model and shape model are proposed to obtain more accurate segmentation results. Wu et al. [4] proposed a method based on Laws' microtexture energies and maximum a posteriori (MAP) estimation to construct a probabilistic deformable model for kidney segmentation. However, the discriminative ability and computational efficiency of its feature selection need to be further improved. In addition, the segmentation results for regions with insignificant gradients are not ideal. Martin-Fernandez et al. [5] used markov random fields and active contours to detect kidney contours from ultrasound images. Although this method provides good topology flexibility, obtaining the most optimized results takes a lot of time. Xie et al. [6] used texture and shape priors to segment the kidney from ultrasound images. This method alleviates the noise interference and improves the segmentation speed, but it cannot solve the effect of blurred boundary. Mendoza et al. [7] segmented KUS images based on active shape models and statistical shape models. However, this method is time-consuming and has poor segmentation results for irregular kidney shapes. Although these semiautomatic segmentation methods have improved the segmentation accuracy of kidneys to varying degrees, they often require manual initialization operations [8]. In addition, these methods still cannot solve the problems caused by intensity energy distribution, heterogeneous structure and variable shapes [1]. Therefore, it is very meaningful to segment the kidney automatically and accurately from the ultrasound image.

Due to the strong nonlinear learning ability of convolutional neural network (CNN), its superiority is fully reflected in the segmentation of natural images [9–13] and medical images [8,14–18]. Recently, CNN technology has also been applied to the segmentation of ultrasound images [19–21]. In 2016, Zhang et al. [22] used two series-connected full convolutional networks (FCN) to segment lymph nodes ultrasound images. Two series-connected FCN models are responsible for coarse and fine lymph node segmentation, respectively. However, the segmentation results of this method on ultrasound images with blurred boundaries are not ideal. In 2017, Wu et al. [23] designed a cascaded FCN to segment prenatal ultrasound (PUS) images. This method alleviates the interference of boundary blur and noise on the segmentation results through the introduction of the auto-context scheme. However, it is difficult to obtain an ideal boundary map for severely cascaded images. In 2018, Kim et al. [24] designed a novel CNN model to segment arterial ultrasound images. In this network, multi-scale input and mixed loss function are introduced. This method uses a multi-label loss function with weighted pixel-wise cross-entropy to alleviate the imbalance of the rate of background, wall and lumen. In the same year, Mishra et al. [25] developed a deep-supervised CNN with attention mechanism to segment ultrasound images. However, this method cannot make full use of the object context information. In 2019, Cunningham et al. [26] designed a DeconvNet to segment neck muscles ultrasound images. In 2020, Shareef et al. [16] designed a small tumor perception network for breast ultrasound (BUS) image segmentation. This method improves the accuracy of small breast tumor segmentation through dense multi-scale convolution. Later, a large number of methods were proposed for the segmentation of BUS images [20,27,28]. Although ultrasound image segmentation based on CNNs has made significant progress. However, the differences between different tissues make their ultrasound images very diverse, and each method cannot obtain the best segmentation results on each type of ultrasound image. Therefore, we need to design a more representative method based on the characteristics of the kidney. In 2020, Yin et al. [8] used the boundary distance map obtained by transfer learning and regression network to segment the kidney. These optimization methods are often limited by the kidney features captured by the transfer learning network [29]. Moreover, it is also a challenge to obtain accurate boundary distance maps from ultrasound images with blurred boundaries and severe cascades. In 2021, Chen et al. [30] proposed an architecture to segment KUS images by fusing structural and texture features. However, capturing complete structural features from ultrasound images with border cascades or incomplete kidneys remains a challenge. Recently, Chen et al. [31] developed a new multi-branch architecture for adequately extracting kidney features from ultrasound images. However, this method ignores the feature reconstruction process, and the segmentation results still have room for improvement.

To more accurately segment the kidney from ultrasound images, we construct a novel multi-scale and deep-supervised CNN model. The main contributions of this paper are as follows:

- A novel multi-scale and deep-supervised network is proposed to segment KUS images. The design of multi-scale inputs pyramid (MSIP) can capture features at different scales to improve the generalization ability of the network. In addition, we develop a multi-output supervision module (MOS) to enable the network to learn to predict more precise segmentation results scale-by-scale.
- Through quantitative analysis on six indicators with several state-of-the-art segmentation methods, our method achieves the best results. The segmentation results show that our method can better reduce the false detection and missed detection of the kidney.

2. Method

Due to the influence of kidney shape, image quality, uneven energy distribution and so on, segmenting kidneys automatically and correctly from ultrasound images is a difficult point [2]. Inspired by CNNs, we specially design a multi-scale and deep-supervised CNN model to segment the kidney from ultrasound images Fig. 2. shows the overall network architecture of the propose method.

F/S/R means Filter/Stride/Rate size; the Output shape is denoted as Height × Width × Channel.

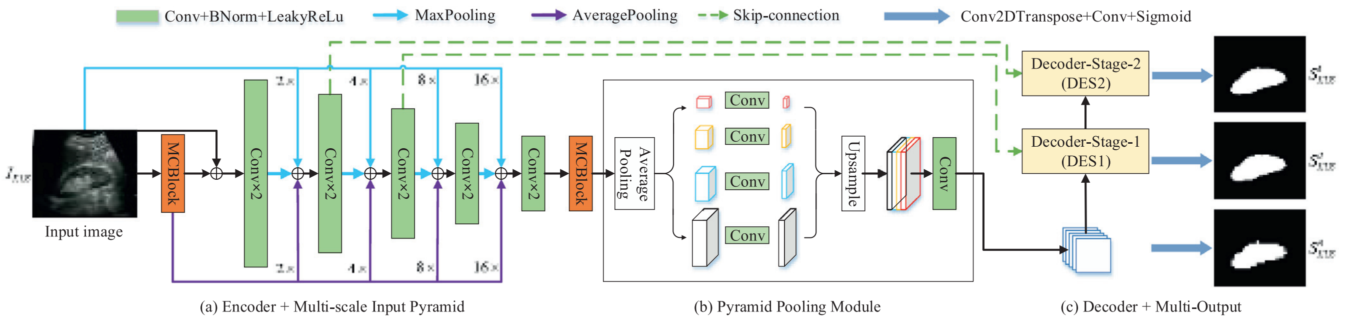

Fig. 2. The overall network architecture of the proposed network. Each decoder-stage (DES) is composed of a Conv2DTranspose and a MCBlock.

Table 1
Detailed Configuration of the Proposed Model.

Encoder			Pyramid Pooling Module			Decoder		
Layer	F/S	Output	Layer	F/S/R	Output	Layer	F/S	Output
MCBlock	–	512 × 512 × 64	AvePooling	32 × 32	1 × 1 × 1024	Multi-Output3	–	512 × 512 × 1
Concat	–	512 × 512 × 65	Conv × 1	1 × 1	1 × 1 × 512	Conv2DTranspose	4 × 4	128 × 128 × 512
Conv × 2	3 × 3	512 × 512 × 64	Upsample	32 × 32	32 × 32 × 512	Concat	–	128 × 128 × 768
MaxPooling	2 × 2	256 × 256 × 64	AvePooling	16 × 16	2 × 2 × 1024	MCBlock	–	128 × 128 × 256
AveragePooling	2 × 2	256 × 256 × 64	Conv × 1	1 × 1	2 × 2 × 512	Multi-Output2	–	512 × 512 × 1
MaxPooling	2 × 2	256 × 256 × 1	Upsample	16 × 16	32 × 32 × 512	Conv2DTranspose	2 × 2	256 × 256 × 256
Concat	–	256 × 256 × 129	AvePooling	8 × 8	4 × 4 × 1024	Concat	–	256 × 256 × 384
Conv × 2	3 × 3	256 × 256 × 128	Conv × 1	1 × 1	4 × 4 × 512	MCBlock	–	256 × 256 × 128
MaxPooling	2 × 2	128 × 128 × 128	Upsample	8 × 8	32 × 32 × 512	Multi-Output1	–	512 × 512 × 1
AveragePooling	4 × 4	128 × 128 × 64	AvePooling	4 × 4	8 × 8 × 1024			
MaxPooling	4 × 4	128 × 128 × 1	Conv × 1	1 × 1	8 × 8 × 512			
Concat	–	128 × 128 × 193	Upsample	4 × 4	32 × 32 × 512			
Conv × 2	3 × 3	128 × 128 × 256	Concat	–	32 × 32 × 2048			
MaxPooling	2 × 2	64 × 64 × 256	Conv × 1	3 × 3	32 × 32 × 512			
AveragePooling	8 × 8	64 × 64 × 64						
MaxPooling	8 × 8	64 × 64 × 1						
Concat	–	64 × 64 × 321						
Conv × 2	3 × 3	64 × 64 × 512						
MaxPooling	2 × 2	32 × 32 × 512						
AveragePooling	16 × 16	32 × 32 × 64						
MaxPooling	16 × 16	32 × 32 × 1						
Concat	–	32 × 32 × 577						
Conv × 2	3 × 3	32 × 32 × 1024						
MCBlock	–	32 × 32 × 1024						

2.1. Network architecture

In Fig. 2 and Table 1, the architecture is described in detail. The network is mainly composed of encoder, pyramid pooling module and decoder. Particularly, we first use MSIP to generates a set of low-level features as the input for the next coding stage. Then the final output of the encoder is processed by the pyramid pooling module to effectively capture multi-scale feature information. Finally, we use different scale features output by the decoder to predict the segmentation results. A novel deep supervision module is included in the decoding stage. By introducing the MOS module, the decoder can predict the segmentation results of the kidney scale-by-scale. The function of the segmentation network can be defined as:

$$S_{KUS} = S_{Net}(I_{KUS}) \qquad (1)$$

where $S_{Net}(\cdot)$ is the function of the method, I_{KUS} is the input image, and S_{KUS} is the segmentation result.

2.2. Multi-scale inputs pyramid (MSIP)

Capturing multi-scale feature information from the input image is an effective method to improve the accuracy of segmentation [32–35]. Abraham et al. constructed an input pyramid by downsampling the input image [32]. However, the single pooling operation of this method easily causes the loss of image feature information. To capture more useful low-level features, we design a novel multi-scale inputs pyramid (MSIP) with parallel branches in the coding stage. As shown in Fig. 3, MSIP contains two parallel branches, which are the max-pooling branch of the input image and the average-pooling branch of low-level features. The MSIP can not only reduce the loss of low-level texture information, but also reduce the sensitivity of the network to input data. There are four down-sampling operations in MSIP and their pool-size is 2, 4, 8, 16.

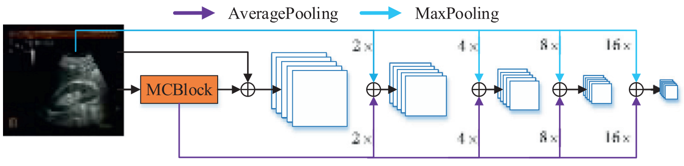

Fig. 3. The description of the proposed Multi-scale inputs pyramid (MSIP).

2.3. Multi-scale convolution block (MCBlock)

The success of CNN benefits from the extraction of target feature information. Studies have shown that the size of the receptive field is very helpful for the representation of target features. In or-

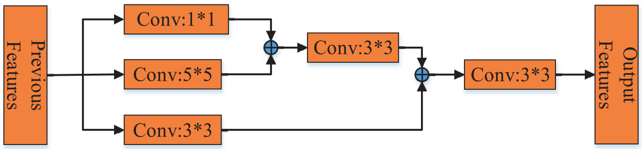

Fig. 4. The detailed description of the proposed MCBlock.

der to obtain feature maps generated by receiving fields of different sizes, we specially design a MCBlock to replace the ordinary convolutional layer. The design of the MCBlock can extract and fuse more refined multi-scale kidney feature information. The structure of the MCBlock is shown in Fig. 4. The processing of input features through MCBlock can be defined as:

$$F_{MC} = S^{C3}\left(S^{C3}\left(S^{C1}\left(F_{Input}\right)+S^{C5}\left(F_{Input}\right)\right)+S^{C3}\left(F_{Input}\right)\right) \tag{2}$$

where $S^{C1}(\cdot)$, $S^{C3}(\cdot)$ and $S^{C5}(\cdot)$ represent the three convolution modules with the convolution kernels size of 1 × 1, 3 × 3, and 5 × 5, respectively.

2.4. Multi-output supervision module (MOS)

The MOS module provides a guarantee for the accurate segmentation of KUS images. Experimental results show that this mechanism can make the boundary of the segmentation result more complete. To achieve the MOS module prediction masks, the multi-channel output of the pooling pyramid and each decoder stage is fed to a convolution layer with kernel size is 1 × 1. Then, the sigmoid layer performs mask prediction. Finally, the mask is subjected to a bilinear up-sampling operation to obtain the input image size. During the training process, given an input image, our segmentation network will output three predicted results. Although every prediction mask is up-sampled to the same size with the input image, the last one has the highest accuracy and hence is taken as the final output of the network. Through the introduction of the MOS module, the prediction mask obtained by the segmentation network can be expressed as:

$$S^i_{KUS} = S_{MOS}(F^i_{DES}), i = 1, 2, 3 \tag{3}$$

where $S_{MOS}(\cdot)$ denote the function of the MOS module, F^i_{DES} represents multi-channel features. S^3_{KUS} is the final output of the network.

3. The description of the experiments

3.1. Datasets

We collected a KUS dataset from 400 patients in two hospitals (Fourth Medical Center of PLA General Hospital and Civil Aviation General Hospital), which contains 400 KUS images. These ultrasound images were acquired by four ultrasound devices (Philips EPIQ7, Mindray, Esaote MyLab, and Hitachi), respectively. The original images acquired by the Philips, Esaote and Hitachi have 768 × 576 pixels, and the original images acquired by the Mindray have 1440 × 1080 pixels. The pixel size of all images ranging from 0.08 mm to 0.32 mm. The regions of interest of the ultrasound images are distributed in the fan, so we crop these KUS images to a size of 512 × 512, as shown in Figs. 1 and 5. During data collection, all identification information is removed to protect the privacy of patients. In the kidney annotation process, we strictly followed the standard annotation procedure. First, the kidney region of each ultrasound image was annotated by a medical student. Then, these preliminary annotation results were revised by an experienced sonographer. Finally, the sonographer-annotated results were further reviewed by a clinically experienced urologist. The disputed annotation area is decided by two doctors after consultation.

The number of patients used for network training and testing was 350 and 50, respectively. In the test dataset, the ultrasound images of some patients are shown in Fig. 7. To improve the generalization ability of the network, we perform data augmentation on the original training images of 350 patients. Our data augmentation methods include rotation (the degrees of rotation are 90°, 180° and 270°), flipping (along the x-axis and y-axis), adding Gaussian noise (mean=0, δ = [5, 10]), adding multiplicative noise (mean=0, δ = [0.05, 0.10]), blurring(the size of blur kernel is 3 × 3) and random gamma transforming (γ = [0.5, 2]). After data augmentation, we obtain 7350 KUS images. 80% of these images are used for network training, and the remaining images are used for network validation. The results of data augmentation on a single image are shown in Fig. 5.

3.2. Implementation and experimental setup

For training, we randomly select 20% from the training dataset as the validation dataset. In this paper, we use Adam algorithm to optimize our network. We set the initial learning rate to 0.001. To prevent the network from overtraining, we will terminate the training process of the network in time according to the loss curve of the validation data. After multiple verifications, the best performance can be achieved when the epoch size and batch size of the network are set to 50 and 8, respectively. We implement our network based on the publicly available framework: TensorFlow 2.2.0. The development environment is Ubuntu 18.04 and python 3.6. An eight-core PC with an Intel i7 7500 U 3.5 GHz CPU and an NVIDIA RTX 3090 GPU.

3.3. Loss function

Through the comparative analysis of the loss function, we finally chose weighted cross-entropy (WCE) loss. The complete loss function of the network can be expressed as:

$$L = \sum_{n=1}^{N} (\ell_{wce})^n \tag{4}$$

where ℓ_{wce} is WCE loss. $(\ell_{wce})^n$ is the loss of the $n - th$ output. N is the number of multi output. As described in Section 2, our segmentation model has three deep supervision outputs, i.e. $N = 3$.

4. Experimental results

4.1. Description of evaluation metrics

In order to highlight the superiority of our method, we use the most popular evaluation indicators for comparative experimental analysis. In particular, we measure the segmentation performance of all segmentation methods on six indicators: Accuracy, Jaccard, Precision, Recall, Dice and Average Symmetric Surface Distance (ASSD). The mathematical expressions of Accuracy, Jaccard, Precision, Recall, Dice and ASSD are shown from Eqs. (5) - (11).

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

$$Jaccard = \frac{TP}{FP + TP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

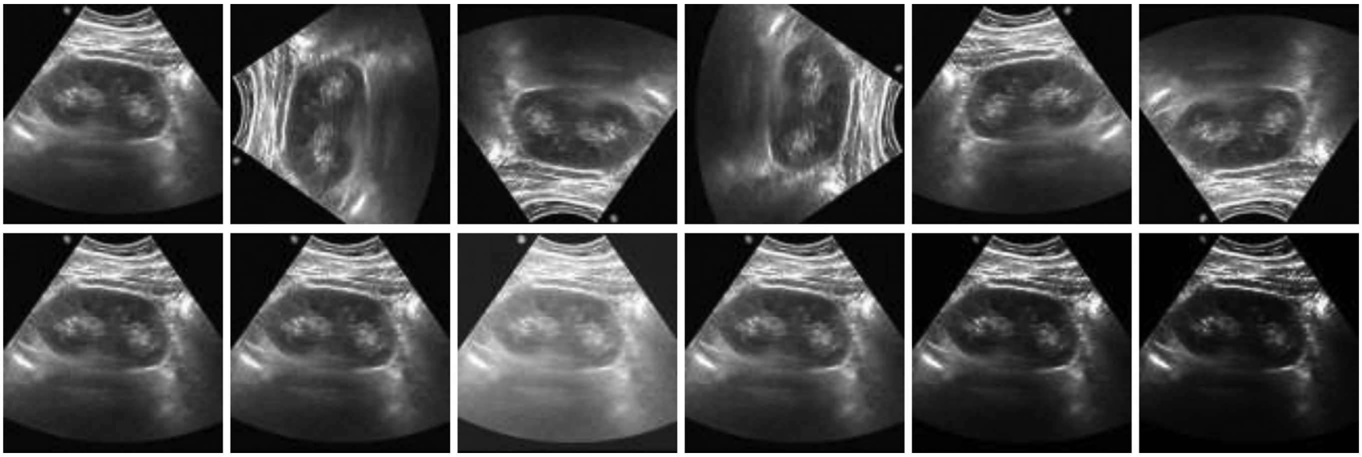

Fig. 5. The part results of data augmentation on a single image.

Table 2
Ablation study on different architectures and loss. The best results are marked with bold texts. The red arrow indicates an increase in results. The green arrow represents a decrease the result.

Ablation	Configurations	Jaccard%	Dice%	Accuracy%	Recall%	Precision%	ASSD
Architecture	Baseline + ℓ_{bce}	87.56	92.72	98.19	91.03	95.68	1.1331
	Baseline + MSIP + ℓ_{bce}	88.36↑	93.41↑	98.33↑	91.40↑	95.81↑	0.9264↑
	Baseline + MSIP + DES2 + ℓ_{bce}	89.15↑	94.07↑	98.42↑	92.73↑	95.88↑	0.7005↑
	Baseline + MSIP + DES3 + ℓ_{bce}	88.98↓	93.86↓	98.34↓	92.05↓	95.83↓	0.7865↓
	Baseline + MSIP + DES2 + MOS + ℓ_{bce}	90.32↑	94.75↑	98.57↑	93.77↑	96.00↑	0.5493↑
Loss	ℓ_{wbce}	**92.18**	**95.86**	**98.86**	**95.47**	**96.38**	**0.350**
	ℓ_{focal}	89.08	94.06	98.40	92.07	**96.38**	0.6302
	$\ell_{bce} + \ell_{dice}$	88.55	93.36	98.38	91.80	95.99	0.8143
	$\ell_{wbce} + \ell_{dice}$	89.80	94.37	98.55	93.05	96.19	0.8893
	$\ell_{focal}+\ell_{dice}$	88.12	93.37	98.25	91.02	96.21	0.9536

DESi means that the network contains i decoding stages, and the baseline network contains one decoding stage.

$$Dice = 2 * \frac{Precision \cdot Recall}{Precision + Recall} \tag{9}$$

$$ASSD = \frac{(mean_{g \in BG}(\min_{s \in BS} d(g,s)) + mean_{s \in BS}(\min_{g \in BG} d(g,s)))}{2} \tag{10}$$

where TP is defined as the positive output. FP is defined as the positive output. TN means the negative output. FN means the negative output. BG denotes boundary voxels of the segmentation G_{KUS}, BS denotes boundary voxels of the segmentation S_{KUS}, $d\ (\cdot,\ \cdot)$ is the Euclidian distance between two points.

4.2. Ablation analysis of the our method

To further analyze the advantages of each network components, we conducted ablation experiments on the network structure. In addition, we analyzed the impact of loss function on network performance. The ablation experiments are mainly conducted on our collected dataset. To evaluate the strengths and weaknesses of each component more fairly, all the segmentation results are not post-processed.

Architecture ablation: we show the effectiveness of the principal components of our network, i.e., multi-scale inputs pyramid (MSIP), decoder stage number (DES), multi-output supervision module (MOS). The baseline network (i.e., first row of Table 2) is constructed by removing MSIP, DES2 and MOS components from our network Table 2. illustrates the results of the architecture ablation study. As we can see, our method achieves the best performance among these configurations. From Table 2, we can see that adding MSIP, DES2 and MOS components to the baseline network can gradually improve the performance of the network. This proves that the MSIP, DES2 and MOS components we designed are working. Through the comparison of Baseline and MSIP (i.e., Baseline + MSIP), we can see that learning features of different scales from KUS images can improve the accuracy of kidney segmentation. It can be concluded from the results of MSIP, DES2 (i.e., Baseline + MSIP + DES2) and DES3 (i.e., Baseline + MSIP + DES3) that appropriately increasing the depth of the decoder can improve the performance of the network. This also proves that the deeper the network is not the better for a particular problem. This is exactly the purpose of our analysis of DES components. From the results of MOS (i.e., Baseline + MSIP + DES2 + MOS), we can see that the design of the multi-output deep supervised module can further improve the segmentation performance of the network. Through the above analysis of the components, we believe that applying them to the existing segmentation network can improve their performance.

Loss ablation: At present, the influence of loss function on the accuracy of ultrasound image segmentation is still unclear. Therefore, we conducted a quantitative analysis of several commonly used segmentation losses (such as binary cross-entropy loss (ℓ_{bce}), weight binary cross-entropy loss (ℓ_{wbce}) and focal loss (ℓ_{focal})). The experimental results show that the weighted cross-entropy loss achieves the best results. In order to further verify the effectiveness of the weighted cross-entropy, we conducted a comparative analysis with the typical mixed loss (i.e., last row of Table 2). In addition, we also constructed two new mixed losses for comparative analysis (i.e., the ninth and tenth rows of Table 2). Experimental results show that more complex loss functions do not bring better segmentation results. By comparing the segmentation results of $\ell_{bce} + \ell_{dice}$, $\ell_{wbce} + \ell_{dice}$, and $\ell_{focal}+\ell_{dice}$, the mixed loss com-

Table 3
Evaluation results of various methods. The best results are marked with bold texts.

	Accuracy (%) Mean± std	Dice (%) Mean± std	Jaccard (%) Mean± std	Precision (%) Mean± std	Recall (%) Mean± std	ASSD Mean± std
U-net	97.04± 2.36	88.39± 9.77	80.37± 13.78	91.17± 9.69	87.20± 13.06	1.5675± 2.69
Att U-net	97.34± 2.49	89.52± 10.23	82.30± 14.16	94.10± 6.51	86.79± 14.60	1.4769± 2.60
FCNN	97.70± 1.71	90.39± 10.81	83.73± 13.31	91.95± 7.77	89.91± 13.05	0.9236± 2.58
Abraham et al.	97.72± 213	91.75± 7.21	85.45± 10.74	94.12± 6.09	90.32± 10.52	0.8299± 1.57
STAN	98.02± 1.63	91.35± 13.86	85.87± 14.74	94.48± 6.25	90.64± 14.69	2.6579± 13.98
SegNet	98.32± 1.58	92.21± 14.38	87.50± 25.25	93.59± 15.01	91.29± 14.85	1.7069± 7.00
PSPNet	98.40± 1.31	93.70± 5.87	88.63± 9.08	96.01± 4.75	92.01± 8.60	0.7832± 1.64
DeeplabV3+	98.41± 1.53	94.05± 5.37	89.16± 8.19	96.25± 5.52	92.18± 6.86	0.6393± 1.20
Ours	**98.86± 0.89**	**95.86± 2.76**	**92.18± 4.85**	**96.38± 3.01**	**95.47± 4.02**	**0.3510± 0.67**

posed of ℓ_{wbce} still achieves the best results. This further illustrates the robustness of weight binary cross-entropy loss to KUS images segmentation.

4.3. Comparision with the state-of-the-arts

Currently, FCNN [9], DeeplabV3+ [10], SegNet [12], PSPNet [13] have been widely used in comparative experiments of image segmentation. In this paper, we also chose these four methods for comparative experiments. In addition, we chose four neural network models for medical image segmentation, which are STAN [16], Abraham et al. [32], U-net [36] and Att U-net [37]. STAN is an advanced BUS images segmentation method. In order to ensure that each comparison method can exert the best performance, we first use their existing network weights to initialize, and then retrain on our training dataset. The evaluation indicators and segmentation results of various methods are in Table 3 and Fig. 7.

As shown in Table 3, our method achieves the best results. DeeplabV3+ and PSPNet achieve the second and third result. Compared with DeeplabV3+, our method increases the Accuracy, Dice, Jaccard, Precision and Recall indexes by 0.46%, 1.93%, 3.39%, 0.14% and 3.57%, respectively. At the same time, the ASSD indicator is reduced by 0.2883. Compared with PSPNet, the advantages of our method are further demonstrated. Accuracy, Dice, Jaccard, Precision and Recall indicators are improved by 0.47%, 2.31%, 4.01%, 0.39% and 3.76%, respectively. The difference on the ASSD indicator is 0.4322. In general, PSPNet and DeeplabV3+ have achieved competitive results but there is still room for improvement. SegNet and STAN are ranked fourth and fifth respectively, which shows that they are not very adaptable to KUS image segmentation. The ranking of the remaining methods is Abraham et al., FCNN, Att U-net and U-net. This shows that their ability to segment KUS images is not strong. The method of Abraham et al. achieves the fourth result on the ASSD indicator, but the results of other indicators are poor. The results of U-net show that it has a poor generalization ability for KUS images segmentation.

In order to further evaluate the segmentation of the comparison method on the KUS image, we have also drawn precision-recall (P-R) and F-measure curves in Fig. 6. The P-R curve represents the confidence level of predicting true positive and false positive classes. The F-measure curve indicates the confidence level of correct prediction of a method. As shown in Fig. 6, our method has achieved the most superior performance compared to other methods, which shows that our method is more suitable for segmentation of KUS images.

Fig. 7 shows the segmentation results of nine methods on the test images. The segmentation results of all methods are not post-processed. From these KUS images, it can be seen that there are obvious noises, severe disease interference, blurred kidney boundaries and heterogeneous structures. It can be seen from the segmentation results that our method is significantly better than other methods and has a lower false detection rate and missed detection rate. The segmentation results of various methods on the effusion interference ultrasound images have relatively small differences. The effusion seriously affected the segmentation results of the kidney, as shown in the first and second rows images. Fortunately, our method can overcome the interference of effusion to a certain extent. Compared with effusion, cyst has less disturbance to the accuracy of kidney segmentation, as shown in the third to sixth rows images. Only the kidneys with larger cysts have obvious false detection. Compared with other methods, our method is hardly interfered by cyst. In general, our method has relatively complete segmentation results for kidneys with disease interference. According to the segmentation results in Fig. 7, our method is better than other methods in the segmentation of kidney images with blurred boundary. As shown in the seventh and ninth rows of Fig. 7, our method still has good results in segmentation of kidney images with fan-shaped shadows. From a visual point of view, FCNN has the worst segmentation results, with obvious blocks on the edges. It can be seen from the segmentation results of U-net that there are obvious false detection and missed detection. In terms of visual effects, U-net, Att U-net and Abraham et al. have achieved similar results, all of which have serious missed detections and false detections. It is worth noting that FCNN has a lower false detection rate than U-net and Att U-net. The false detection of STAN has been greatly

Input Label FCNN U-net Att U-net Abraham et al. STAN SegNet PSPNet DeeplabV3+ Ours alleviated, but there are still serious false detections in individual images, as shown in the third, fourth and sixth rows images. The segmentation results of SegNet, PSPNet and DeeplabV3+ are relatively good. However, these methods have poor segmentation results for kidneys with uneven intensity distribution, as shown in the fourth row of Fig. 7. In addition, these three methods still have some false detection and missed detection.

It is well known that ultrasound images have severe speckle noise interference. How to accurately model the degradation process of ultrasonic images has always been a research hotspot. However, due to the difference between the simulated noise degradation model and the real noise degradation model, it is difficult to quantitatively evaluate the impact of real noise degradation on the segmentation accuracy of ultrasound images. Nonetheless, the segmentation results on clinical data acquired by four ultrasound devices by our method can show that our method is not sensitive to real noise interference. In addition to noise interference, the variable kidney morphology also easily interferes with the segmentation accuracy of the network. Compared with state-of-the-art segmentation methods, the proposed method is less perturbed by kidney morphology as shown in Fig. 7.

Through the ablation experiment analysis of the network components, it is proved that every component we design is working. Compared with advanced segmentation methods, our method achieves the best results on six popular evaluation indicators. In the analysis of segmentation results, although our method has cer-

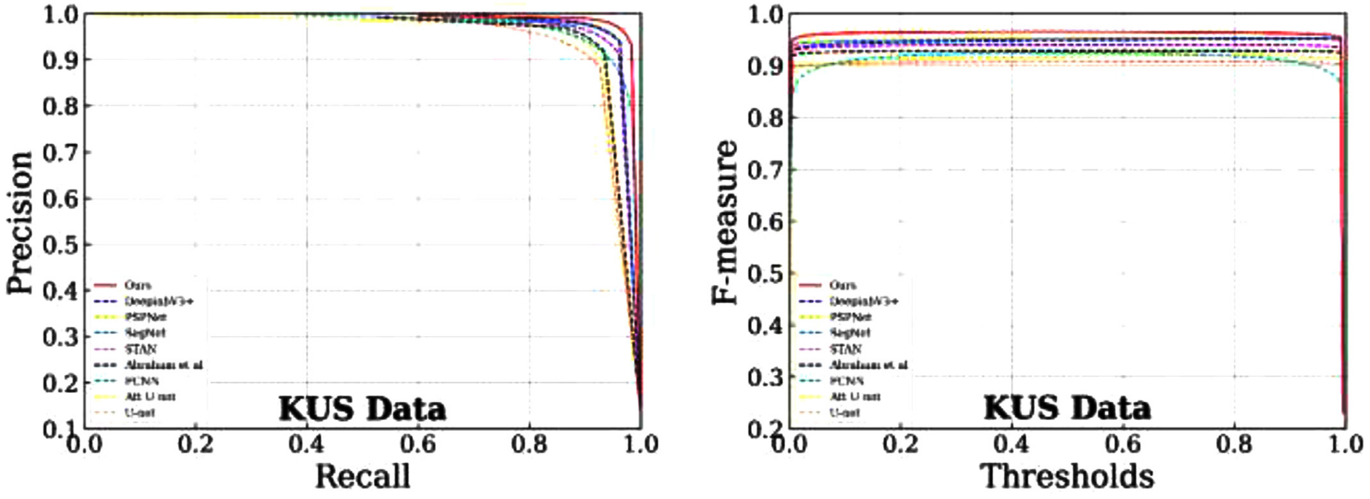

Fig. 6. Illustration of P-R curves (the left) and F-measure curves (the right) on the comparison segmentation method.

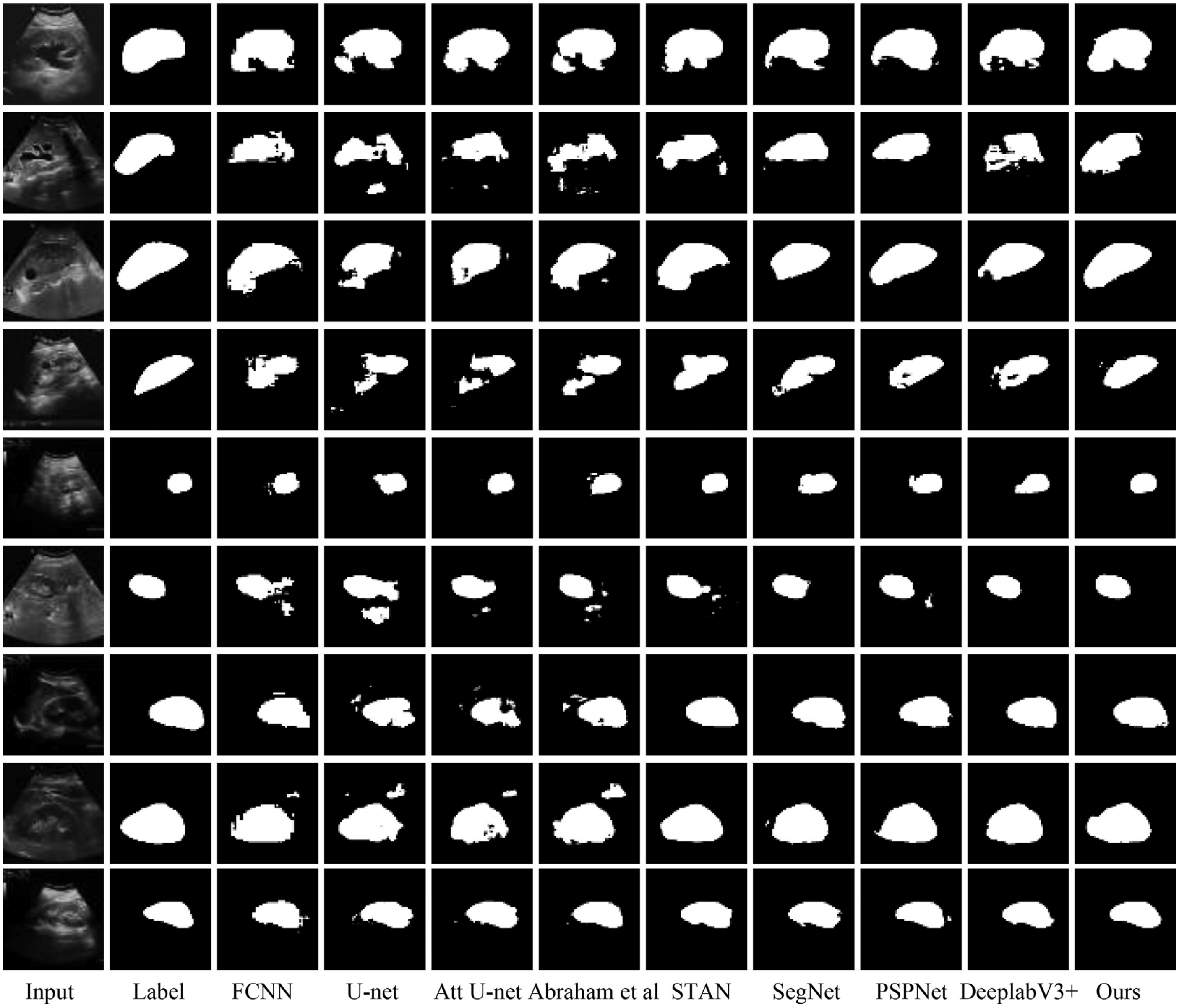

Fig. 7. The segmentation results of KUS images by different methods. These segmentation results include disease interference images, boundary blurred images and ultrasound shadow images.

tain false detection and missed detection, it still achieves the best results compared with other methods. Specially, our method has better robustness to KUS images and is less disturbed by various factors. In summary, the segmentation method proposed in this paper alleviates the problem of automatic segmentation of KUS images to a certain extent.

5. Conclusion

Inspired by CNN technology, we construct a multi-scale and deep-supervised network to segment kidney accurately from ultrasound images. Specifically, the architecture includes encoder, pyramid pooling module and decoder, which are in charge of extracting low-level kidney feature, capturing multi-scale kidney feature and reconstructing high-level kidney feature, respectively. In the encoder, we design a multi-scale input pyramid (MSIP) with parallel branches to capture features at different scales. In the decoder, the introduce of the multi-output supervision module (MOS) can make the network learn to predict the segmentation results scale-by-scale. In addition, we design a multi-scale convolution block (MCBlock) in the network to further extract and fuse more refined multi-scale image information. The ablation experiment analysis of the network proves that these components can gradually improve the performance of the network. The effectiveness of our method is further verified by the comparative experimental analysis with the advanced segmentation method. In addition, the comparative analysis with existing methods is very helpful for our future work.

Although our method has achieved good results, there are still some shortcomings that need to be overcome. First, due to the influence of image quality, blurred boundaries and heterogeneous structure, our method still has some false detections and missed detections. Second, the real-time nature of ultrasound images is not considered, which is very important for clinical practice. In the future, we will conduct in-depth research on the effects of ultrasound image quality, blurred boundaries and heterogeneous structures on the ultrasound image segmentation results. In addition, the time cost will be further considered when constructing the network model.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Acknowledgment

This work is supported by the National Key R&D Program of China (Grant No. 2017YFB1302803) and National Natural Science Foundation of China (Grant No: U1913207, 51875394).

References

[1] H.R. Torres, S. Queiros, P. Morais, B. Oliveira, J.C. Fonseca, J.L. Vilaca, Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: a systematic review, Comput. Methods Programs Biomed. 157 (2018) 49–67, doi:10.1016/j.cmpb.2018.01.014.

[2] J.A. Noble, D. Boukerroui, Ultrasound image segmentation: a survey, IEEE Trans. Med. Imaging 25 (2006) 987–1010, doi:10.1109/tmi.2006.877092.

[3] Q. Zheng, S. Warner, G. Tasian, Y. Fan, A dynamic graph cuts method with integrated multiple feature maps for segmenting kidneys in 2D ultrasound images, Acad. Radiol. 25 (2018) 1136–1145, doi:10.1016/j.acra.2018.01.004.

[4] C.H. Wu, Y.N. Sun, Segmentation of kidney from ultrasound B-mode images with texture-based classification, Comput. Methods Programs Biomed. 84 (2006) 114–123, doi:10.1016/j.cmpb.2006.09.009.

[5] M. Martín-Fernández, C. Alberola-López, An approach for contour detection of human kidneys from ultrasound images using Markov random fields and active contours, Med. Image Anal. 9 (2005) 1–23, doi:10.1016/j.media.2004.05.001.

[6] J. Xie, Y. Jiang, H.T. Tsui, Segmentation of kidney from ultrasound images based on texture and shape priors, IEEE Trans. Med. Imaging 24 (2005) 45–57, doi:10.1109/tmi.2004.837792.

[7] C.S. Mendoza, X. Kang, N. Safdar, E. Myers, C.A. Peters, M.G. Linguraru, W. Dc, Kidney segmentation in ultrasound via genetic initialization and Active Shape Models with rotation correction, 2013 IEEE 10th, Int. Symp. Biomed. Imaging. (2013) 69–72.

[8] S. Yin, Q. Peng, H. Li, Z. Zhang, X. You, K. Fischer, S.L. Furth, G.E. Tasian, Y. Fan, Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks, Med. Image Anal. 60 (2020) 101602, doi:10.1016/j.media.2019.101602.

[9] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2015, pp. 3431–3440.

[10] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proc. Eur. Conf. Comput. Vis, 2018, pp. 801–818.

[11] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 834–848, doi:10.1109/TPAMI.2017.2699184.

[12] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 2481–2495, doi:10.1109/TPAMI.2016.2644615.

[13] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2017, pp. 2881–2890.

[14] S. Yin, Z. Zhang, H. Li, Q. Peng, X. You, S. Furth, G. Tasian, Y. Fan, Fully-automatic segmentation of kidneys in clinical ultrasound images using a boundary distance regression network, in: 2019 IEEE 16th Int. Symp. Biomed. Imaging (ISBI 2019), IEEE, 2019: pp. 1741–1744.

[15] B. Shareef, A. Vakanski, M. Xian, P.E. Freer, ESTAN: enhanced Small Tumor-Aware Network for Breast Ultrasound Image Segmentation, arXiv preprint arXiv:2009.12894. (2020).

[16] B. Shareef, M. Xian, A. Vakanski, Stan: small tumor-aware network for breast ultrasound image segmentation, 2020 IEEE 17th Int. Symp. Biomed. Imaging, IEEE (2020) 1–5.

[17] J.M.J. Valanarasu, V.A. Sindagi, I. Hacihaliloglu, V.M. Patel, Kiu-net: overcomplete convolutional architectures for biomedical image and volumetric segmentation, arXiv preprint arXiv:2010.01663. (2020).

[18] H. Lee, J. Park, J.Y. Hwang, Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image, IEEE Trans. Ultrason. Ferroelectr. Freq. Control. 67 (2020) 1344–1353.

[19] B. Behboodi, H. Rivaz, Ultrasound Segmentation Using U-Net: Learning from Simulated Data and Testing On Real Data, IEEE. (2019).

[20] B. Behboodi, M. Amiri, R. Brooks, H.R.B.T.-2020 I. 17th I.S. on B.I. (ISBI), Breast lesion segmentation in ultrasound images with limited annotated data, in: 2020.

[21] B. Behboodi, M. Fortin, C.J. Belasso, R. Brooks, H.B.T.-2020 42nd A.I.C. of the I.E. in M. and B.S. (EMBC) in conjunction with the 43rd A.C. of the C.M. and B.E.S. Rivaz, Receptive Field Size As a Key Design Parameter for Ultrasound Image Segmentation with U-Net, in: 2020.

[22] Y. Zhang, M. Ying, Y. Lin, A.T. Ahuja, D.Z.B.T.-I.I.C. on B. & B. Chen, Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images, in: 2016.

[23] L. Wu, X. Yang, S. Li, T. Wang, N.B.T.-2017 I. 14th I.S. on B.I. (ISBI 2017) dong, cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation, in: 2017.

[24] S. Kim, Y. Jang, B. Jeon, Y. Hong, H. Chang, Fully automatic segmentation of coronary arteries based on deep neural network in intravascular ultrasound images, intravascular imaging and computer assisted stenting and large-scale annotation of biomedical data and expert label synthesis, 2018.

[25] Mishra Deepak, Chaudhury Santanu, Sarkar Mukul, Singh Arvinder, Soin, Ultrasound Image Segmentation: a Deeply Supervised Network with Attention to Boundaries, IEEE Trans. Bio Med. Eng. (2018).

[26] R. Cunningham, M.B. Sánchez, I.D. Loram, Ultrasound segmentation of cervical muscle during head motion: a dataset and a benchmark using deconvolutional neural networks, (2019).

[27] C. Xue, L. Zhu, H. Fu, X. Hu, X. Li, H. Zhang, P. Heng, Global guidance network for breast lesion segmentation in ultrasound images, Med. Image Anal. 70 (2021) 101989.

[28] W.K. Moon, Y.-.W. Lee, H.-.H. Ke, S.H. Lee, C.-.S. Huang, R.-.F. Chang, Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks, Comput. Methods Programs Biomed. 190 (2020) 105361.

[29] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S.X. Li, D. Ni, T. Wang, Deep learning in medical ultrasound analysis: a review, ngineering 5 (2019) 261–275, doi:10.1016/j.eng.2018.11.020.

[30] G. Chen, Y. Dai, R. Li, Y. Zhao, L. Cui, X. Yin, SDFNet: automatic segmentation of kidney ultrasound images using multi-scale low-level structural feature, Expert Syst. Appl. 185 (2021) 115619, doi:10.1016/j.eswa.2021.115619.

[31] G. Chen, Y. Dai, J. Zhang, X. Yin, L. Cui, MBANet: multi-branch aware network for kidney ultrasound images segmentation, Comput. Biol. Med. 141 (2021) 105140, doi:10.1016/j.compbiomed.2021.105140.

[32] N. Abraham, N.M.B.T.-2019 I. 16th I.S. on B.I. (ISBI) Khan, A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation, in: 2019.

[33] G. Chen, Z. Gao, Q. Wang, Q. Luo, U-net like deep autoencoders for deblurring atmospheric turbulence, J. Electron. Imaging. 28 (2019) 53024.

[34] G. Chen, Z. Gao, Q. Wang, Q. Luo, Blind de-convolution of images degraded by atmospheric turbulence, Appl. Soft Comput. 89 (2020) 106131.

[35] G. Chen, Z. Gao, P. Zhu, Z. Chen, Learning a Multi-scale Deep Residual Network of Dilated-Convolution for Image Denoising, in: 2020 IEEE 5th Int. Conf. Cloud Comput. Big Data Anal., IEEE, 2020: pp. 348–353.

[36] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Int. Conf. Med. Image Comput. Comput. Interv., Springer, 2015: pp. 234–241.

[37] O. Oktay, J. Schlemper, L.Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-Net: learning Where to Look for the Pancreas, (2018). http://arxiv.org/abs/1804.03999.