W-Net: Two-stage segmentation for multi-center kidney ultrasound

Yu-Chi Chang
Graduate Institute of Library Information and Archival Studies
National Chengchi University
Taipei, Taiwan
yc.chang@g.nccu.edu.tw

Chung-Ming Lo
Graduate Institute of Library Information and Archival Studies
National Chengchi University
Taipei, Taiwan
buddylo@nccu.edu.tw

Yi-Kong Chen
Division of Nephrology, Department of Internal Medicine
Kaohsiung Medical University Hospital
Kaohsiung, Taiwan
k75456@gmail.com

Ping-Hsun Wu
Division of Nephrology, Department of Internal Medicine
Kaohsiung Medical University Hospital
Kaohsiung, Taiwan
970392KMUH@gmail.com

Hsing Luh
Department of Mathematical Sciences
National Chengchi University
Taipei, Taiwan
slu@nccu.edu.tw

***Abstract*—The global death rate of chronic kidney disease (CKD) continues to increase and becomes a serious health issue. Ultrasound imaging is significant in the evaluation of CKD. However, there is a challenge posed by quality differences in multi-center datasets for kidney ultrasound image segmentation. Confronting the problem, this study applied the W-Net based on the double U-Net architecture which was respectively trained in two stages. In the first stage, the pixel-wise nnU-Net was pretrained by 4586 images and fine-tuned by 534 images. In the second stage, the region-wise nnU-Net was trained from the inference of the first stage by 72 images and achieved a 6.95% improvement from the first stage. It can bring more evidence about the practical application of deep learning-based segmentation in kidney ultrasound and its potential use in clinics.**

Keywords*—*W-Net, kidney, ultrasound, segmentation, multi-center

I. Introduction

The global burden of CKD is substantial and continues to increase. Approximately 10% of the adult population worldwide is affected by various forms of CKD, leading to an estimated 12 million deaths annually[1]. The age-standardized mortality rate attributable to CKD saw an increase of 41.5% from 1990 to 2017 [1]. Ultrasound imaging is a key step in the evaluation of CKD. The advantages of ultrasound include nonionizing and real-time, making it suitable for kidney examination [2]. The findings in kidney properties such as length, width, cortical thickness, and the margin shape are particularly crucial in the progression of kidney diseases[2]. Therefore, to identify accurate kidney status for effective disease monitoring and diagnosis, image segmentation techniques were employed in the literature. Upon automatic segmentation, the monitoring and diagnosis of kidney status can be more efficiency and objective[3, 4].

However, the appearances of ultrasonic images may vary substantially due to the various settings between different manufactures, models, and even scanning protocols. A well-designed segmentation method should have substantial generalized ability to be used in clinic. Therefore, collecting multi-center datasets to encompass various image appearances in brightness, contrast, and speckle noise for the evaluation of segmentation methods is necessary. In this study, a total of three different datasets collected from the United States, Canada, and Taiwan with the amounts of 4586, 534, and 72 were used for the model training and testing. Based on these datasets, W-Net, a two-stage segmentation method was proposed to reduce the effect of variabilities among different image appearances and qualities. The result can bring more evidence about the practical value of applying deep learning-based segmentation in kidney ultrasound and its potential use in clinic.

II. Material and Methods

The W-Net involves two-stages of nnU-Net [5] segmentation like the concatenation of two u-shape. The adoption of nnU-Net due to its dice similarity coefficient (DSC=0.9806) better than TransUnet (DSC=0.9651) [6], FTransCNN (DSC=0.9697) [7], and HA-SNET (DSC=0.9701) [8] in the prior experiment using CT2US and UBC datasets. In this study, the W-Net was trained by three different datasets for multi-center kidney ultrasound segmentation. In the first stage, pixel-wise nnU-net, was initially trained by CT2US dataset [9] consisting of 4,586 images and fine-tuned by UBC dataset [10] consisting of 534 images. In the second stage, region-wise nnU-net, trained from the inference based on fine-tune by Kaohsiung Medical University Hospital (KMUH) dataset containing 72 ultrasound images. Based on the second stage of training, the input images were the result generated from the first stage and thus only region-wise information can be learned to solve the variations between pixels from different ultrasound sources.

A. Dataset

Three datasets were used to train the W-Net segmentation model. CT2US, was generated through CycleGAN [11] from MICCAI's KITS19 dataset having 210 patients. The results were 4,586 ultrasound images with masks. UBC, was from a single urban tertiary hospital in Canada. The images were acquired from patients with CKD, prospective kidney donors, and those with a transplanted kidney between January 2015 and September 2019 and had 534 ultrasound images with masks. KMUH dataset was collected from KMUH in November 2023 including images of polycystic kidney and renal cyst. The contours of these 72 images were manually delineated by nephrologists.

B. Pixel-wise segmentation

The first stage of W-Net was used to classify pixels into the kidney area and background tissues using nnU-Net [12]. In this stage, nnU-Net was proposed to do pixel-wise segmentation by training on the CT2US dataset first and followed by the subsequent fine-tuning on the UBC dataset. The network was trained on five folds with a learning rate of

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00274

0.01 annealed throughout training, using the SGD optimizer. The loss function used in training was a combination of dice and cross-entropy loss as below:

$$\mathcal{L}_P = \mathcal{L}_{dice} + \mathcal{L}_{CE} \quad (1)$$

$$\mathcal{L}_{Dice}(y, \hat{y}) = 1 - \frac{2y\hat{y}}{y+\hat{y}+1} \quad (2)$$

$$\mathcal{L}_{CE}(y, \hat{y}) = -(ylog(\hat{y}) + (1-y)\log(1-\hat{y})) \quad (3)$$

Here, $\hat{y}$ is the predicted value by the prediction model. Consider an original kidney ultrasound image $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, where $H \times W$ represents the size of each image and C denotes the RGB channel. When $\mathbf{X}$ is input into the pixel-wise nnU-Net, $\mathbf{X_p} \in \{0, 1\}^{H \times W}$ is derived. $\mathbf{X_p}$ consists of N regions x_i such as:

$$\mathbf{X_p} = x_1 \cup x_2 \cup \ldots \cup x_N \quad (4)$$

C. Region-wise segmentation

In the second stage of W-Net, another region-wise nnU-Net was trained from the inference of the first stage by the dataset from KMUH. The input images were inferenced masks rather than ultrasonic pixels. Compared to the ground truth, some areas having similar appearances to kidney tissues were misclassified. Therefore, the right part of W-Net focused on regional information which was absent in the first stage such as shapes, boundaries, and locations. Additionally, the deeper network layers were also helpful. The network was trained on five folds with a learning rate=0.01, optimizer=SGD, and the loss function in (1).

Consider $\mathbf{X_p}$ as the input of the region-wise nnU-Net, then the output $\mathbf{X_R} \in \{0, 1\}^{H \times W}$ is derived. $\mathbf{X_R}$ is one of the candidate region x_i which is close to the actual renal image:

$$\mathbf{X_R} = candidate(x_i), i = 1, \ldots, N \quad (5)$$

III. Results

The first stage of W-Net has shown the best performance on the UBC dataset in terms of the DSC=0.9488 when CT2US dataset was applied on pretrain. the CT2US dataset was adopted due to its popularity in the literature. However, when applying the first stage of W-Net to KMUH dataset for inference, the DSC was only 0.8258 due to the quality difference. To address this issue, the second stage of W-Net was subsequently applied to extract features of shapes, numbers, and location information which were absent in the first stage. The second stage of W-Net achieved a better performance on KMUH dataset in terms of DSC=0.8832. That is, W-Net showed an improvement of 6.95% over the conventional nnU-Net. TABLE I shows that W-Net has better performance than U-Net framework on both stages. Figure 1 shows the final segmentation results obtained from W-Net applied to KMUH dataset. W-Net demonstrates superior segmentation compared to the popular U-Net. Additionally, the experiment shows using two stage was helpful no matter what network was used. The limitation of this study is the lack of wider range comparisons. Only the most popular U-Net, nnU-Net were compared with W-Net in the two stage experiment.

TABLE I. Segmentation Results of DSC in Two Stage of W-Net and U-Net on KMUH Dataset

Network	1st stage	2nd stage
U-Net [5]	0.7346	0.7762
W-Net	0.8258	0.8832

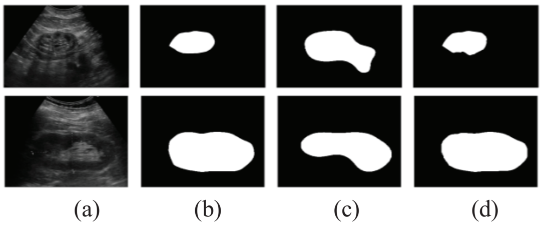

Fig. 1. Comparisions between W-Net and U-Net. (a) original image (b) ground truth (c) U-Net (DSC=0.6051 (up), 0.8474 (down) (d) W-Net (DSC=0.9740 (up) and 0.9896 (down)).

IV. Conclusion

This study presents the W-Net, a two-stage segmentation network for the automatic delineation of kidney. By reducing the variances among multi-center datasets, the DSC was improved from 0.8258 to 0.8832 as an improvement of 6.95%. W-Net has the practical applicability in clinics.

References

[1] B. Bikbov *et al.*, "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *The lancet,* vol. 395, no. 10225, pp. 709-733, 2020.

[2] R. K. Singla, M. Kadatz, R. Rohling, and C. Nguan, "Kidney ultrasound for nephrologists: a review," *Kidney Medicine,* vol. 4, no. 6, p. 100464, 2022.

[3] C.-C. Kuo *et al.*, "Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning," *NPJ digital medicine,* vol. 2, no. 1, p. 29, 2019.

[4] S. Xun *et al.*, "Current Status, Prospect and Bottleneck of Ultrasound AI Development: A Systemic Review," *Advanced Ultrasound in Diagnosis & Therapy (AUDT),* vol. 7, no. 2, 2023.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015: Springer, pp. 234-241.

[6] J. Chen *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306,* 2021.

[7] W. Ding *et al.*, "FTransCNN: Fusing Transformer and a CNN based on fuzzy logic for uncertain medical image segmentation," *Information Fusion,* p. 101880, 2023.

[8] X. Pan, J. Ai, and J. Zhang, "HA-SNet: A Best Choice to Solve Ultrasonic Images Based on Hybrid Attention Network," in *2023 3rd International Conference on Computer, Control and Robotics (ICCCR)*, 2023: IEEE, pp. 106-110.

[9] Y. Song, J. Zheng, L. Lei, Z. Ni, B. Zhao, and Y. Hu, "CT2US: Cross-modal transfer learning for kidney segmentation in ultrasound images with synthesized data," *Ultrasonics,* vol. 122, p. 106706, 2022.

[10] R. Singla *et al.*, "The open kidney ultrasound data set," in *International Workshop on Advances in Simplifying Medical Ultrasound*, 2023: Springer, pp. 155-164.

[11] L. Wang, W. Chen, W. Yang, F. Bi, and F. R. Yu, "A state-of-the-art review on image synthesis with generative adversarial networks," *IEEE Access,* vol. 8, pp. 63514-63537, 2020.

[12] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods,* vol. 18, no. 2, pp. 203-211, 2021.