IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 May 2023, accepted 5 June 2023, date of publication 8 June 2023, date of current version 15 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3284029

RESEARCH ARTICLE

Automatic Segmentation of Kidney Volume Using Multi-Module Hybrid Based U-Shape in Polycystic Kidney Disease

HAOYANG CUI[1,*], YIYI MA[2,*], MING YANG[2,*], YANG LU[1], MINGZI ZHANG[1], LILI FU[2], CHICHENG FU[1], BEILIN SU[2], CHUAN HE[1], CHENG XUE[2], CHANGLIN MEI[2], AND SHUWEI SONG[2]

[1]Shanghai Aitrox Technology Company Ltd., Shanghai 200050, China
[2]Department of Nephrology, Shanghai Changzheng Hospital, Naval Medical University, Shanghai 200433, China

Corresponding authors: Shuwei Song (songsw@smmu.edu.cn) and Changlin Mei (changlinmei@smmu.edu.cn)

*Haoyang Cui, Yiyi Ma, and Ming Yang contributed equally to this work.

This work was supported in part by the Shanghai Municipal Key Clinical Specialty under Grant shslczdzk02503, in part by the National Natural Science Foundation of China under Grant 81873595, and in part by the Special Clinical Research Project of Shanghai Municipal Health Commission under Grant 202040241.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Shanghai Changzheng Hospital under Application No. CZEC(2013)-12.

ABSTRACT Polycystic kidney disease (Autosomal Dominant Polycystic Kidney Disease, ADPKD) is the most common genetic disease of the kidney, and the measurement of Total Kidney Volume (TKV) in clinical research of this disease is essential to study the progression of ADPKD. At present, the volume segmentation of polycystic kidneys mainly relies on doctors to manually outline the kidney boundary on the radiological image. This process is time-consuming, labor-intensive, inefficient, subjective, and difficult to guarantee consistency. In the research of this paper, A multi-module hybrid U-shape segmentation method is proposed (HUNet), which introduces wavelet pooling, cascade residual, and efficient multi-head self-attention into the U-shape structure. We use wavelet pooling instead of traditional down-sampling to reduce the loss of detailed features, the use of cascaded residual modules can improve the ability of model feature reuse, and the use of efficient multi-head self-attention modules can effectively capture global multi-scale information. In the decoding process of the U-shape, the corresponding loss value of each decoder will be calculated, and finally, the total loss value of the model will be obtained by weighted average. The method was trained and tested on the polycystic kidney dataset provided by Shanghai Changzheng Hospital. We automatically segmented the ADPKD in MRI images using the proposed method with a remarkably high Dice similarity coefficient relative to the manual segmentation (mean=0.915). The percentage difference between the total kidney volume values using manual and HUNet methods was only 0.4%. The proposed approach enables fast and accurate TKV measurement.

INDEX TERMS Polycystic kidney, wavelet pooling, cascade residual, multi-head self-attention.

I. INTRODUCTION

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is the most common inherited disorder of the kidneys. It is characterized by the enlargement of the kidneys caused by the progressive development of renal cysts. It is one of the leading causes of end-stage renal disease (ESRD) [1]. Results from previous studies have demonstrated an association between total kidney volume (TKV) and renal function [2], and researchers can use total kidney volume (TKV) as a measure for early diagnosis and prognostic assessment [3], [4]. With the development of Magnetic Resonance Imaging (MRI),

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Oliva.

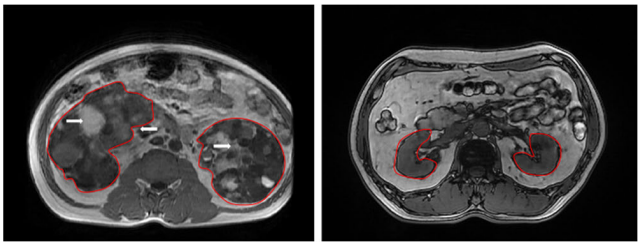

FIGURE 1. Compared with healthy kidneys (right panel), ADPKD kidneys (left panel) are difficult to segment due to severe morphological changes. White arrows show surface cysts of different sizes and irregular surface contours.

Computed Tomography (CT), and other medical imaging technologies, high-resolution images of the renal region can be acquired non-invasively layer by layer. Clinicians perform manual segmentation of kidney region images to segment kidney regions. However, there are inefficiencies, poor consistency, and subjective experience in the way clinicians manually segment kidney regions layer by layer. It is therefore crucial to develop rapid and reliable methods for TKV quantification. Polycystic kidneys are characterized by their markedly irregular shape and size in comparison to normal kidneys and sometimes surface irregularities are prominent due to the presence of surface cysts of different sizes. In polycystic kidney disease, as illustrated in Figure 1, there are numerous and large cysts which may even rupture and bleed, resulting in more image layers and irregular image contours, thereby making outlining more challenging. For clinicians without a background in polycystic kidney disease research, it is necessary to undergo map recognition training and software operation, which can be both time-consuming and labor-intensive.

Therefore, developing fully automatic segmentation methods for fast and accurate TKV estimation remains a challenging problem. At present, with the development of artificial intelligence, especially deep learning has made remarkable achievements in the field of image recognition and semantic segmentation [5], [6], which has promoted the application of deep learning in the field of medical image segmentation.

II. RELATED WORK

In recent years, deep learning has been widely used in the field of medical image segmentation, and many researchers have made corresponding research and contributions in kidney segmentation based on radiological images. Traditional segmentation methods usually refer to the use of prior knowledge and imaging information to achieve target segmentation. Pohle and Toennies [7] proposed an adaptive region growing algorithm, which automatically learns its homogeneity criteria according to the characteristics of the kidney region, and designs region growth criteria based on the selected seed points. Among semi-automatic approaches to MRI, Daum et al. [8] used 3D random walks while Racimora et al. [9]. proposed active contours and morphological operations for the segmentation of polycystic kidneys. Sharma et al. [10] proposed a random forest and geodesic distance volume-based method for 3D segmentation of polycystic kidneys in ADPKD patients with severe renal insufficiency. Traditional methods generally require human intervention. They cannot achieve fully automated segmentation, and the processing process is cumbersome, while the consistency is also poor.

At present, the segmentation methods based on deep learning mainly adopt a data-driven approach, and its performance is closely related to the quantity and quality of data. The model is trained by constructing a loss function so that the model has the ability to efficiently extract image features and can automatically segment the target area. In recent years, Convolutional Neural Networks (CNN) have shown excellent performance in computer vision tasks such as image classification, object detection, and semantic segmentation. The main advantage of CNNs over many other machines learning based methods (e.g., Random Forests [11], SVM [12]) is that they do not require handcrafted features. Initially, Long et al. [6] proposed the Full Convolutional Network (FCN), which replaces all fully connected layers with convolutional layers to achieve pixel-by-pixel prediction without fixing the image size, with good generalization performance and high segmentation accuracy without any post-processing. The first fully convolutional neural network with an encoder and decoder structure was proposed by Badrinarayanan et al. [13]. Ronneberger et al. [14] proposed the classic U-net structure, which uses a skip connection to fuse the output features of the encoder and decoder and achieved good results in medical image segmentation. Drawing on the network structure of the first 10 layers of VGG, Sharma et al. [10] designed a fully convolutional neural network for polycystic kidney segmentation. However, due to cystic lesions, the shape of the kidney has changed significantly, and the kidney region cannot be well located. Although CNN methods have strong learning capabilities in various medical image segmentation tasks, the locality of convolutional layers in CNNs, limits the capability of learning long-range spatial dependencies. Transformer architecture using the self-attention mechanism has been successful in natural language processing (NLP) [15], with its capability of capturing long-range dependency. Recently, multiple methods were proposed that explore the possibility of using transformer-based models for the task of 2D image segmentation [16], [17], [18]. Goel et al. [19] used a U-Net architecture with an EfficientNet encoder for accurate segmentation of polycystic kidneys on MRI, which reduced the time required for expert contouring. Raj et al. [20] used a deep learning model based on attention mechanism that can automatically segment the kidney images, and used a new loss function called the cosine loss function that can effectively address the class imbalance problem in deep learning models and improve their performance. Kim et al. [21] proposed a deep learning model that incorporates a U-Net architecture with residual connections to achieve better segmentation accuracy. Additionally, they introduced a loss function called the Dice-Sørensen coefficient, which is more effective in

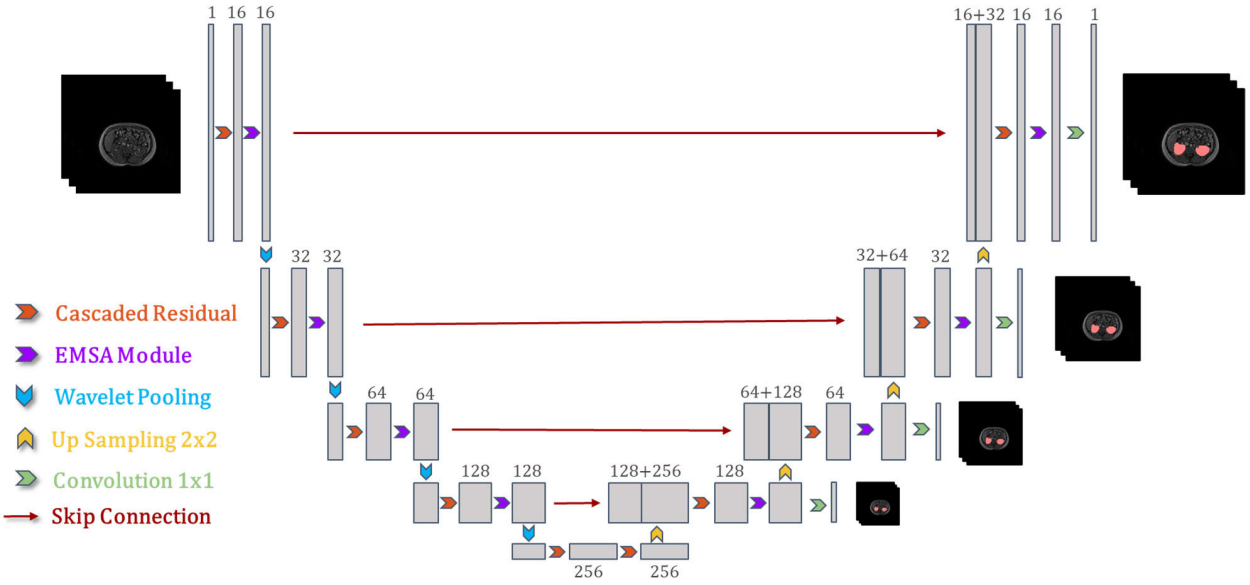

FIGURE 2. The hybrid architecture of the proposed HUNet.

handling class imbalance and boundary misalignments in segmentation tasks. The proposed deep learning model was trained and tested on a dataset of CT scans from patients with ADPKD.

In this work, we present a multi-module hybrid U-shape model (HUNet), a simple yet powerful hybrid architecture for the automatic segmentation of polycystic kidneys, trained end-to-end, on slice-wise axial-MRI sections. The innovations of the HUNet method are as follows: **a)** The multi-scale analysis of wavelet transform [22] is introduced into the U-shape structure, and the use of wavelet pooling replaces the traditional Max/Mean pooling, thereby reducing the information loss of feature maps during the pooling process. **b)** We design a cascaded residual module to avoid network degradation while improving the model's ability to reuse features. **c)** Towards enhanced quality of segmentation, we seek to apply self-attention to extract detailed long-range relationships on high-resolution feature maps. **d)** In this paper, the HUNet segmentation model is constructed in a multi-module cascade way, which can realize the plug-and-play of each module.

In addition, in the decoding stage of HUNet, the output result of each decoder is fused as the final segmentation result of the model using weighted average. The experimental results show that the model proposed in this paper can effectively segment polycystic kidneys and has a high consistency with the segmentation results of doctors. Given the design of HUNet, our framework is expected to generalize well to other medical image segmentation.

III. METHODS

A. NETWORK ARCHITECTURE BASED ON U-SHAPE STRUCTURE

Fig.2 highlights the architecture of HUNet, we use wavelet pooling, cascaded residuals, and efficient multi-head self-attention as the basic modules of U-shape to construct a HUNet for polycystic kidney segmentation. The encoder pathway is similar to the typical classification network to extract more high-level semantic feature layer by layer. Then the decoder pathway recovers the localization for every voxel and utilizes the feature information to classify it. These segment outputs would be compared with corresponding resolution labels and then used to calculate the final loss function. Such supervision encourages the network to predict correctly from the low-resolution feature maps which will be up-sampled to be full-resolution feature maps.

B. WAVELET POOLING & CASCADE RESIDUALS

The two most popular forms of pooling are max pooling and average pooling. Max pooling involves taking the maximum value of a region and selecting it for compressing the feature map. Average pooling involves computing the average of a region and selecting it for a compressed feature map. While max and average pooling both are effective and simple methods, they also have drawbacks. Depending on the data, max pooling can remove details in images. This happens if the main details have less intensity than the insignificant details. Furthermore, max pooling commonly overfits training data [23], [24]. Depending on the data, average pooling can dilute relevant details in the image. The averaging of data with values much lower than significant details causes this action [23], [24]. Figure 3 illustrates these shortcomings using the toy image example:

The classic fast discrete wavelet transform (DWT) [25] is an efficient implementation of the two-dimensional discrete wavelet transform. Because wavelet transform has the advantages of fast, simple, and no redundant information after transformation, it is widely used in the field of image processing [26], [27]. Two-dimensional wavelet

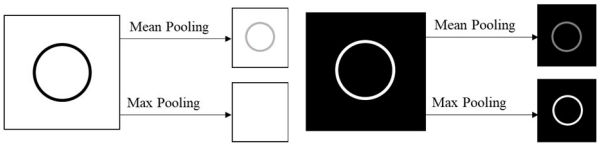

FIGURE 3. Shortcomings of Max & Average Pooling.

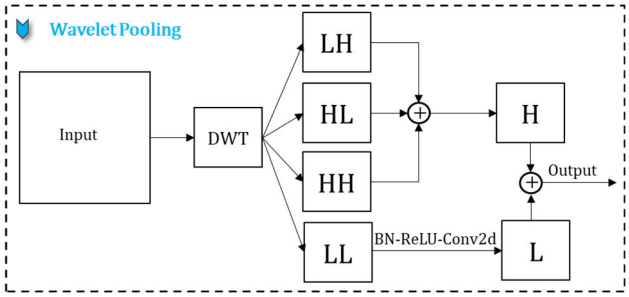

FIGURE 4. Schematic diagram of wavelet pooling.

decomposition will down-sample the image along with the row and column directions, so after a scale transformation, the image size becomes a quarter of the original, thus achieving the same effect as traditional pooling. Since the wavelet decomposition preserves the high-frequency components lost in the low-frequency components, it can make up for the detailed features lost in the pooling process. Based on this theory, wavelet pooling is used to replace the traditional Max/Mean pooling, and the high-frequency components of wavelet decomposition are also introduced into the U-shape structure, so as to make up for the detailed features lost in the pooling process. In this paper, we only use one-level wavelet decomposition for the feature map, so as to obtain the low-frequency and high-frequency information of the feature map. We take the low-frequency components as the pooling result of the feature map and add three high-frequency components to the next layer by means of skip connections. The wavelet pooling module is shown in Figure 4.

where LL denotes the low-frequency component, which is an approximation of the image, HL denotes the high-frequency horizontal component of the image, LH denotes the high-frequency vertical component of the image, and HH denotes the high-frequency diagonal component of the image.

The wavelet pooling has a simple structure and can be introduced into any CNN network architecture to achieve a plug-and-play effect. It is mainly divided into a low-frequency part and a high-frequency part, and the low-frequency part obtains L through a convolution operation. The values of the three high-frequency parts are added to obtain H. Finally, add the values of L and H again, and keep the number of channels unchanged to get the final output result.

In addition, this paper also improves the traditional residual module [28] and designs a cascade residual module. In cascaded residuals, 1×1 convolution is first used to perform cross-channel information fusion, followed by an identity mapping to achieve direct transfer of intra-layer

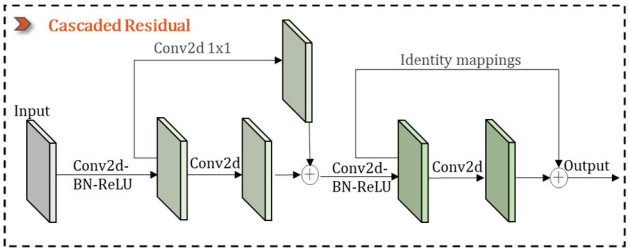

FIGURE 5. Cascaded Residual Modules.

information, to improve the ability of model feature reuse, as shown in Figure 5.

C. EFFICIENT MULTI-HEAD SELF-ATTENTION MODULE (EMSA)

The Transformer is built upon the multi-head self-attention module [15], which allows the model to jointly infer attention from different representation subspaces. The results from multiple heads are concatenated and then transformed with a feed-forward network. Since images are highly structured data, most pixels in high-resolution feature maps within local footprints have similar features, with the exception of boundary regions. Therefore, the pair-wise attention computation among all pixels is highly inefficient and redundant. We design an efficient multi-head self-attention mechanism (EMSA) for learning long-range dependency, as shown in Figure 6. The key idea is to apply convolutional layers to extract local intensity features to avoid large-scale pretraining of attention mechanisms while using multi-head self-attention to capture long-range association information.

The input feature map of EMSA is $X \in \mathbb{R}^{C\times H\times W}$, where H, W are the spatial height, width and C is the number of channels. Three 1×1 convolutions are used to encode X to query, key, and value embeddings: $Q, K, V \in \mathbb{R}^{c\times H\times W}$, where c is the dimension of embedding in each head. In order to reduce the computational complexity, a convolutional layer with kernel size $= s \times s$ and stride $= s$ adopted for reducing to encode dimension.

The Q, K, V is then reshaped to $Q \in \mathbb{R}^{HW\times c}$, $K \in \mathbb{R}^{\frac{HW}{s^2}\times c}$ and $V \in \mathbb{R}^{\frac{HW}{s^2}\times c}$. We use two downsampled convolutional layers to encode K and V into low-dimensional embedding.

Multi-head self-attention is constructed by multiple independent self-attention, and the output results of each self-attention are connected on channels to achieve the effect of integration. The proposed efficient multi-head self-attention is now:

$$Attention\,(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{c}}\right)V \tag{1}$$

$$EMSA = Conv\,(Concat\,(head_1; head_2, \ldots, head_h)) \tag{2}$$

$$head_i = Attention\,(Q, K, V) \tag{3}$$

By doing this, we reduce the computational complexity from $O(n^2c)$ to $O(nkc)$, where $n = (H \times W)$, $k = (H \times W)/s^2$, and s is the sampling step size. The paper selects $s = 32$.

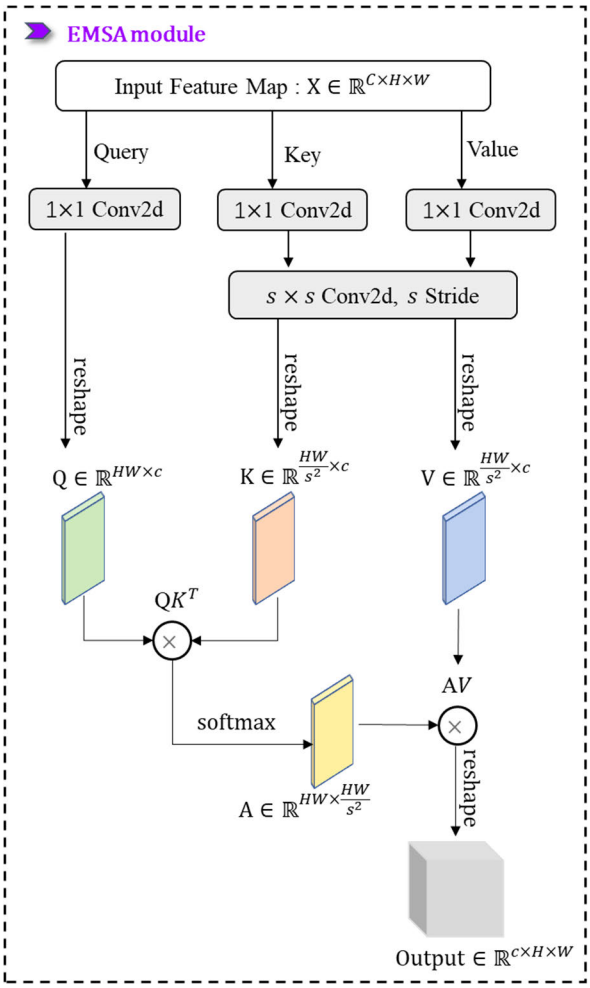

FIGURE 6. The proposed efficient multi-head self-attention (EMSA).

D. LOSS FUNCTION

Due to the serious category imbalance problem in medical imaging data, this paper adopts the Dice coefficient (Dice Similarity Coefficient, DSC) [29] as the loss function of the model, so as to overcome the imbalance problem of positive and negative samples and accelerate the convergence of the model. The calculation formula is as follows.

$$\mathcal{L}_{Dice} = \frac{2\left|V_{seg} \cap V_{gt}\right|}{\left|V_{seg}\right| + \left|V_{gt}\right|} \tag{4}$$

where V_{seg} denotes the predicted segmentation result and V_{gt} denotes the doctor-labeled segmentation result.

In order to constrain the boundary of the segmentation area, we introduce the 95% Hausdorff Distance(HD) Loss [30], the calculation formula is as follows. The 95% HD uses the 95th percentile of the distances between ground truth and prediction surface point sets. As a result, the impact of a very small subset of outliers is minimized when calculating HD.

$$\mathcal{L}\left(X_{seg}, Y_{gt}\right)_{HD} = \max\left[\max_{x \in X_{seg}} \min_{y \in Y_{gt}} \|x - y\|, \max_{y \in Y_{gt}} \min_{x \in X_{seg}} \|y - x\|\right] \tag{5}$$

TABLE 1. Dataset distribution.

	Train	Validation	Test	Total
Patients	355	101	50	506
MRI Slices	26667	7403	3808	37878

where X_{seg} *and* Y_{gt} denote prediction surface and ground truth point sets.

The total loss function is:

$$\mathcal{L}_{seg} = \lambda \mathcal{L}_{Dice} + (1 - \lambda)\mathcal{L}_{HD} \tag{6}$$

where λ is weighting parameters to balance the different loss terms.

In the decoding stage of HUNet, the 4 segment outputs would be compared with corresponding resolution labels and then used to calculate the final loss function. Finally, the weighted average method is used to obtain the total loss value of the entire network. The calculation formula is as follows.

$$\mathcal{L}_{total} = \mathcal{L}1_{seg} + CosineDecay\left(\mathcal{L}2_{seg} + \mathcal{L}3_{seg} + \mathcal{L}4_{seg}\right) \tag{7}$$

$$CosineDecay = \frac{1}{2}\left(1 + \cos\left(\frac{t\pi}{T}\right)\right)\alpha \tag{8}$$

where the *CosineDecay* controls the second term weight, t is the current step, T is the number of steps in the entire training, α is the initial weight.

IV. EXPERIMENT AND RESULTS

A. EXPERIMENTAL DATA

The experimental data in this paper were collected from MRI images of clinical cases in Shanghai Changzheng Hospital. The polycystic kidney region was manually outlined by two experienced radiologists to determine the gold standard for segmentation. The male-to-female ratio is 251:255, with an age range between 15 and 72 years old and an average of 41 years old. The data is in DICOM format, the pixel spacing is 1.875mm, the slice spacing is 4mm, and the MRI image size is 256 × 256×N (N is the number of slices in the MRI sequence), a total of 506 cases of polycystic kidney MRI sequences. The data distribution is shown in Table 1. We perform head-to-tail closure and morphological operations on the outlined contours to generate mask of the target region. The MRI image and mask of the polycystic kidney have the same matrix and voxel size. In order to enable the HUNet network to focus on the learning of the target region as well as to reduce memory consumption, we eliminate the slice without the kidney region in the MRI sequence. The preprocessed images and mask were finally used for HUNet training and testing.

To mitigate overfitting and improve generalization, we augment the data by translation, rotation, and horizontal flipping, use adaptive histogram equalization to enhance image contrast, and normalize the pixel value range to [0, 1]. Furthermore, slices were randomly shuffled before feeding to the CNN. Using a sampling size of 32 (batch-size) for each iteration.

B. EXPERIMENTAL SETUP

The experimental configuration is as follows: training was performed on a workstation with an Intel(R) Xeon(R) Gold 5118 CPU@2.30GHz and 8 NVIDIA GeForce RTX 2080Ti, system is Ubuntu 20.04.1 TLS. Development tools for Python and PyTorch deep learning frameworks. The ranges of Dice loss and Hausdorff Distance (HD) loss were calculated individually. Dice loss values were generally between 0.0 and 1.0, whereas HD loss values typically fell within the range of 0.0 to 4.0. Our objective is twofold: to scale the losses within reasonable ranges and to ensure that HD loss values surpass those of Dice loss. Consequently, this enables additional constraints on the contours of the segmentation results during the later stages of training. The loss weights λ and α were set to 0.8 and 1. Each layer was updated using error back-propagation with adaptive moment estimation optimizer (ADAM) [31], which is a stochastic optimization technique. The exponential decay rates for the moment estimates B1 and B2 are 0.9 and 0.999 respectively, with an epsilon of 10e-8. The learning rate for determining to what extent the newly acquired information overrides the old information was initially 0.0002. We initialized the weights in the encoder and the decoder layers using the kaiming initialization [32]. We maintain the same learning rate for the first 100 epochs and decay the learning rate linearly to zero for the next 100 epochs.

C. EVALUATION

In order to comprehensively evaluate the segmentation accuracy of the model in this paper, we use the Dice coefficient and the Jaccard coefficient as the main evaluation indicators, which are defined as follows.

$$Dice = \frac{2\left|V_{seg} \cap V_{gt}\right|}{\left|V_{seg}\right| + \left|V_{gt}\right|} \tag{9}$$

$$Jaccard = \frac{\left|V_{seg} \cap V_{gt}\right|}{\left|V_{seg} \cup V_{gt}\right|} \tag{10}$$

where V_{seg} denotes the predicted segmentation result and V_{gt} denotes the doctor-labeled segmentation result.

Since image segmentation can be regarded as a pixel-level classification task, we also use Precision and Recall to further evaluate the segmentation accuracy of the model, calculated as follows.

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

where TP represents: the number of positive samples that are correctly identified as positive samples; TN represents: the number of negative samples that are correctly identified as negative samples; FP represents: the number of negative samples that are incorrectly identified as positive samples; FN represents: the number of positive samples that are wrong Number of negative samples identified. The higher the above evaluation indicators, the better the segmentation effect of the model.

TABLE 2. Comparison results of different pooling methods.

Basis Function	Accuracy	Precision	Recall	F1-score
Max	0.9049	0.9019	0.8970	0.8994
Average	0.9172	0.8978	0.9043	0.9010
Haar	0.9273	0.9387	0.9284	0.9335
Daubechies	0.9216	0.9343	0.8913	0.9123
Biorthogonal	**0.9295**	**0.9746**	**0.9359**	**0.9548**

D. TOTAL KIDNEY VOLUME COMPUTATION

All MRI datasets were manually segmented by clinical experts and trained personnel to obtain ground-truth annotations for kidneys. We then performed a morphological closing operation to recover potential holes within predicted kidney regions and to remove any small isolated noise pixels wrongly predicted as foreground pixels. Finally, TKV is calculated as the number of foreground pixels multiplied by the pixel spacing in x and y direction and the corresponding slice thickness.

E. BASIS FUNCTIONS FOR WAVELET POOLING

Different wavelet basic functions will get different low and high-frequency components after wavelet decomposition. In order to select the optimal wavelet basis function, this paper compares and verifies the classic Harr, Biorthogonal and Daubechies wavelet basis functions. Therefore, we built a simple five-layer CNN network to perform binary classification tasks on the CIFAR [33] public dataset to compare the performance of different wavelet basis functions, as shown in Figure 7. Figure 7(b) contains five convolutional layers and two Max/Average pooling layers. Since a convolution operation is also performed in the Wavelet Pooling (WD), three convolution layers are shown in Figure 6(a), the purpose of which is to ensure that the model parameters of the two are consistent. After calculation, the training parameters of the two models are both 589762.

It can be seen from the various evaluation indicators in Table 2 that compared with the maximum pooling and average pooling, the introduction of the wavelet pooling module in the network can significantly improve the classification performance of the CNN model. From the comparison results of different wavelet basis functions, the use of Biorthogonal wavelet basis functions has the greatest improvement in CNN classification performance.

F. RESULTS

The ablation experiment of the multi-module hybrid is shown in Table 3. Unet was utilized as the baseline, and the segmentation results were compared upon the introduction of various modules. The segmentation results obtained using cascaded residuals were comparable to the baseline results, with Dice and Jaccard indices approximately 0.894 and 0.811, respectively. Introducing wavelet pooling

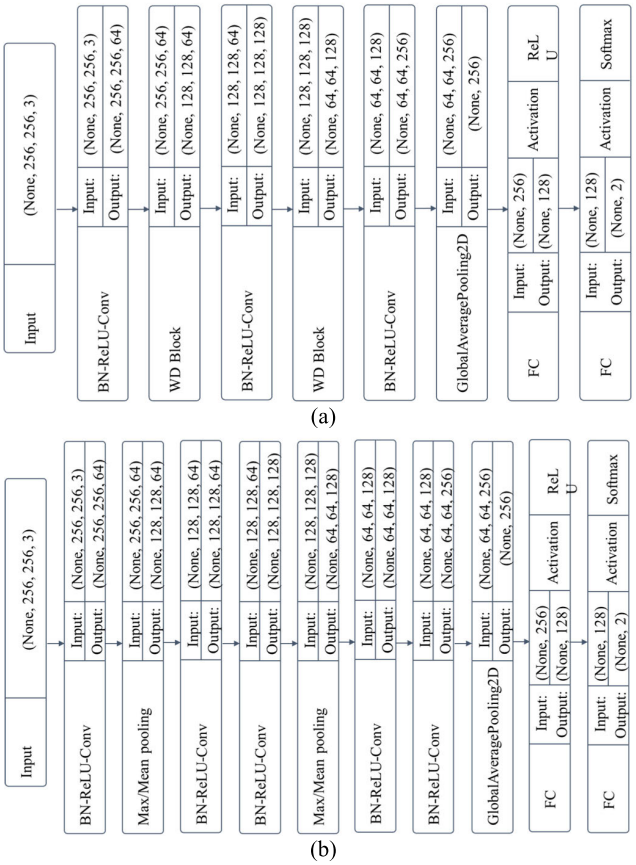

FIGURE 7. Five-layer CNN structure block diagram. (a) CNN model embedded in wavelet pooling block, (b) Traditional CNN model.

TABLE 3. Ablation Study of the Multi-Module Hybrid on the Polycystic Kidney Dataset. The experiment employed the Unet architecture as the baseline and incorporated wavelet pooling and EMSA modules, resulting in noteworthy enhancements in the model's performance. However, the inclusion of cascaded residuals did not yield a substantial improvement in the model's performance. Conversely, the integrated model utilizing all three modules exhibited superior segmentation results on polycystic kidney MRI data.

Architecture	Dice	Jaccard
Baseline Unet	0.894±0.043	0.811±0.065
Wavelet Pooling (WP) + Unet	0.901±0.046	0.825±0.070
Cascade Residuals (CR) + Unet	0.895±0.043	0.810±0.064
EMSA Module + Unet	0.905±0.049	0.830±0.074
WP + CR + EMSA + Unet	**0.915±0.031**	**0.844±0.050**

and EMAS modules separately resulted in substantial performance improvements, with the Dice index surpassing 0.9. Furthermore, the integration of all three modules led to a further enhancement in segmentation quality. The experiment conclusively demonstrated that the integrated HUNet model improved the segmentation performance on polycystic kidney MRI data.

For the test phase, the HUNet model segmented 50 patients in 32.8s, an average of 0.64s per patient, compared to approximately 30 minutes per patient for clinicians to segment manually. We could use the proposed method to automatically segment kidneys in MRI images with a high Dice similarity coefficient relative to manual segmentation (mean±SD = 0.915±0.031 in the HUNet experiment).

Table 4 shows the average segmentation results of different models on the test set. It can be seen that the segmentation performance of HUNet with multi-module hybrid is significantly better than SegNet, Unet and ResUnet. Figure 8 reports the boxplots of the Dice, Jaccard, Precision and Recall metrics for 50 patient cases. From the analysis of Fig.8, one can see that, the proposed HUNet method achieved higher median values and mean values than other models on all four-metrics considered. From the data distribution of the boxplot, HUNet has a smaller variance than other models, indicating that our model is stable for ADPKD segmentation. To assess the statistical significance of these results, we performed T-test for pairwise comparisons. HUNet showed statistically significant differences, in terms of Dice and Jaccard, when compared against the SegNet, Unet and ResUnet ($p < 0.0001$).

In assessing Precision and Recall, we treated the regional segmentation of polycystic kidneys as a pixel-level classification. Since the Precision-Recall (P-R) curve focuses more on positive samples than the ROC curve, the P-R curve can more intuitively reflect the performance of the model on a specific dataset. Therefore, the Average Precision (AP) value of the Area Under the P-R curve is used to evaluate the overall average classification performance of the model on the entire test set, as shown in Figure 9. Compared with other models, HUNet also achieves the best AP value of 0.969.

In order to further intuitively compare the differences between different network models for ADPKD segmentation results, we selected some typical segmentation results for intuitive qualitative analysis and compared them with the gold standard manually marked by experts, as shown in Figure 10. The red curve is the gold standard contour, and the green curve is the contour generated by the deep model. It can be seen that for larger MRI slices in the kidney region (e.g., slice 24), all models have achieved high segmentation accuracy, and the Dice coefficient can reach 0.930 or more, but the segmentation results of HUNet are closer to the gold standard. For MRI slices with small kidney regions and low contrast (e.g., slices 9 and 11), it is difficult for the segmentation model to achieve high segmentation accuracy. From the segmentation results of difficult samples, it can be seen that the HUNet model can also obtain relatively accurate segmentation results when the target region is small and low contrast, while other models are difficult to accurately identify the kidney region. In addition, the size of the left and right kidneys in humans is inconsistent. In slices located at the top or bottom of the kidney, there may be kidney tissue on only one side (e.g., slices 7 and 65), making the model prone to over-segmentation (e.g., ResUnet, Unet, and SegNet), while the HUNet model can accurately localize the kidney and achieve high segmentation accuracy. We counted the Dice value of slices with kidney tissue present on only one side (the number of 397), and compared significant differences using paired t-tests. HUNet showed statistically

TABLE 4. Quantitative comparison of segmentation performance on ADPKD test dataset (mean±SD).

Method	Dice	Jaccard	Precision	Recall
SegNet [13]	0.865±0.042	0.764±0.063	0.896±0.032	0.838±0.062
Unet [14]	0.894±0.043	0.811±0.065	0.892±0.054	0.896±0.036
ResUnet [34]	0.875±0.051	0.781±0.076	0.855±0.078	0.902±0.042
HUNet	**0.915±0.031**	**0.844±0.050**	**0.918±0.034**	**0.913±0.038**

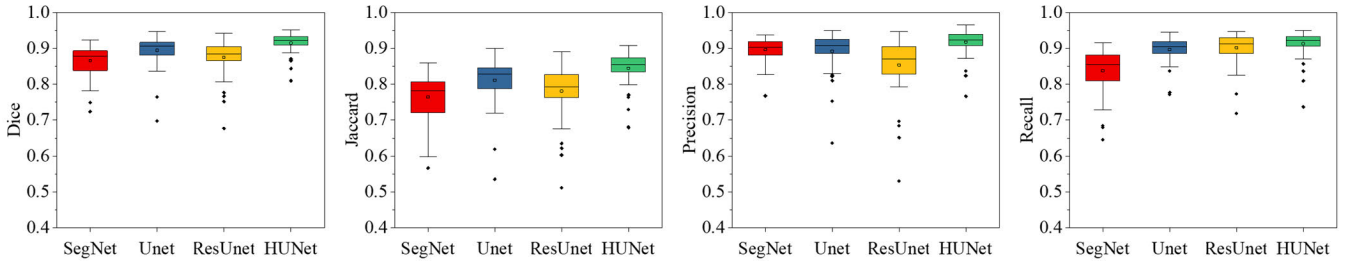

FIGURE 8. The boxplot of the segmentation outputs for the 50 patients in divided test dataset. Boxplots comparing Dice, Jaccard, Precision and Recall metrics for test dataset held out for performance evaluation.

TABLE 5. Paired samples t-test were performed on Dice values for all unilateral kidney slices from the top or bottom of the kidney region.

Pair		N	Variable 1		Variable 2		Paired differences			
			Mean	SD	Mean	SD	Mean	SD	95% Confidence Interval	P (t-test)
HUNet	ResUnet	397	0.7851	0.1928	0.7041	0.2003	-0.08094	0.1893	-0.09962 to -0.06226	<0.0001
HUNet	Unet	397	0.7851	0.1928	0.7438	0.1857	-0.04121	0.1802	-0.05899 to -0.02343	<0.0001
HUNet	SegNet	397	0.7851	0.1928	0.6951	0.1978	-0.09001	0.1904	-0.1088 to -0.07122	<0.0001

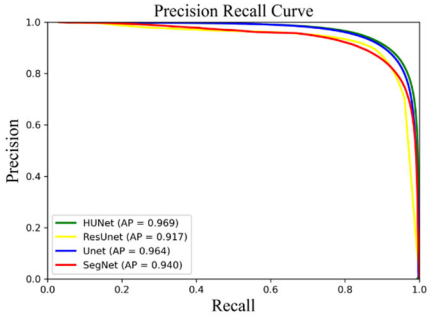

FIGURE 9. Compare the P-R curves and AP values of different models on ADPKD pixel-level classification.

significant differences, in terms of Dice, when compared against the SegNet, Unet and ResUnet, as shown in Table 5.

Comprehensive quantitative and qualitative experimental results show that in the case of the small target area, blurred boundary and low grayscale contrast, other models have different degrees of under-segmentation and over-segmentation, while the HUNet model has better segmentation accuracy.

Reconstruction the three-dimensional shape of the kidney from the MRI sequence can accurately locate the kidney area to obtain more detailed information, which plays an important role in auxiliary diagnosis. Therefore, in this paper, the three-dimensional model of the kidney can be reconstructed from the segmentation results of the MRI sequence, as shown in Figure 11. The segmentation model is used to segment each slice in the polycystic kidney MRI sequence in turn, and then the segmentation results are stacked from top to bottom to reconstruct the three-dimensional model of the polycystic kidney, which provides strong support for subsequent quantitative diagnosis. The visual result of three example kidney subjects is presented in Figure 11. As it is visualized, the proposed HUNet successfully segmented the kidney, which is closer to the manual segmentation results of doctors than other methods. Especially in the top and bottom regions of the kidney, other methods are difficult to achieve accurate segmentation, and our method performs better in segmenting these difficult regions.

Finally, we performed volume measurements on kidney segmentation from CNNs and compared automatic TKV with real TKV in terms of accuracy and precision of measurements. Manual and automated segmentation methods showed comparable performance in evaluation total kidney volume measurements. Bland-Altman plots between total kidney volumes generated using manual and automated segmentation are shown in Figure 12. Bland Altman plots were used to further determine the agreement between the two methods. The Bland-Altman plot shows that the mean value of the percentage difference in total kidney volume measured by the automatic segmentation method proposed by HUNet is closest to 0(mean=0.4%). Compared with other methods, HUNet has higher consistency with manual in measuring total kidney volume.

V. DISCUSSION

TKV is one of the important key indicators to evaluate the severity of the disease in patients with ADPKD and to predict the progression of the disease. The rapid and accurate acquisition of TKV will greatly improve the clinical

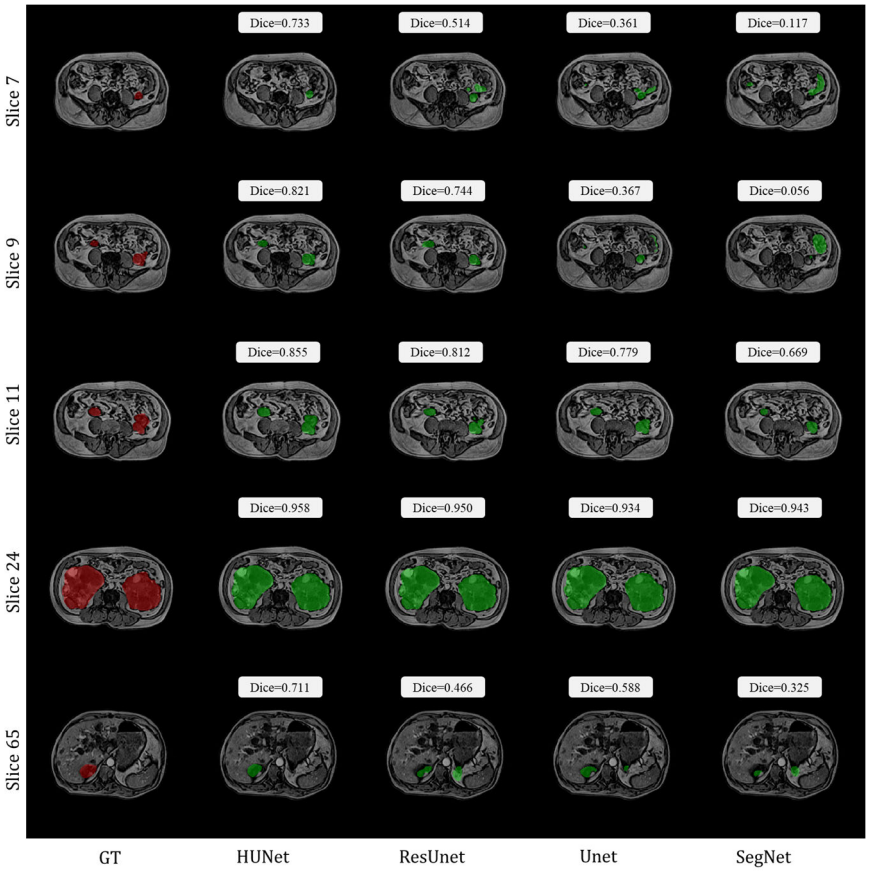

FIGURE 10. Qualitative comparison of different models in the ADPKD test set. Five segmentations (green mask) of ADPKD from different slices of the same MRI sequence are shown. Corresponding physician-annotated kidney segmentation (red mask).

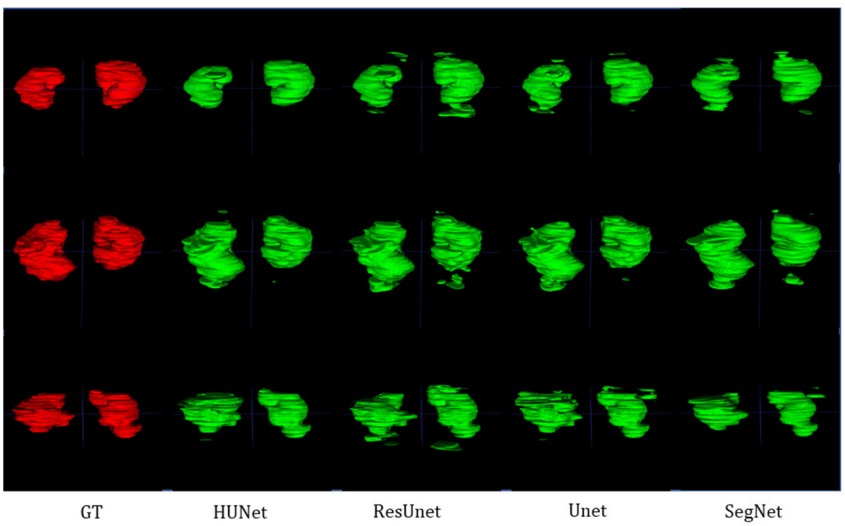

FIGURE 11. Visualizing three-dimensional segmentation results of polycystic kidneys in different segmentation models.

diagnosis and treatment status of ADPKD patients. This study introduces a fully automated segmentation artificial deep neural network for kidney volume assessment. It may provide an alternative to laborious, dedicated, and expensive manual tracing, which is performed in our clinical trials to assess kidney volume and needs high skill currently.

This paper presented the application of U-shaped network-based network in polycystic kidney volume measurement. The polycystic kidney data set provided by Shanghai

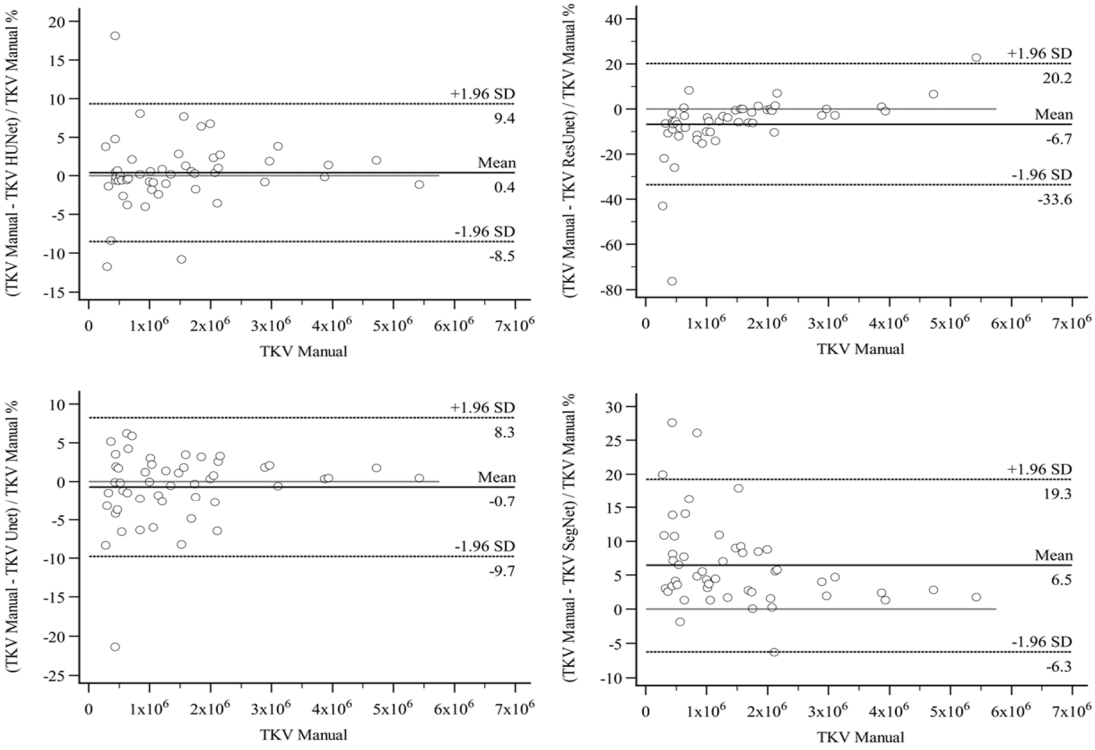

FIGURE 12. Bland-Altman plots between total kidney volumes, generated using manual and deep learning.

Changzheng Hospital was used for training and testing, and the proposed model was compared with other models with similar structures. As can be seen from Table 3, the segmentation performance of HUNet with multi-module hybrid is significantly better than SegNet, Unet and ResUnet. The reason may be that SegNet achieves target segmentation based on a simple encoding and decoding path and does not fuse the underlying feature information; On the basis of SegNet, Unet uses skip connection to fuse the feature information in the encoder, but the feature information between layers is not fully reused, resulting in a learning bottleneck in the model. ResUnet uses a residual module in the layer to overcome the learning bottleneck. However, on the one hand, these models have the phenomenon of feature loss during their respective pooling processes, and on the other hand, they may be limited by the network receptive field, making them too concerned about local information and difficult to effectively learn global information, resulting in difficulty in correctly segmenting the kidneys with kidneys similar background region.

The HUNet proposed in this paper introduces wavelet pooling into the U-shape structure to replace the traditional pooling layer, and improves the residual module, which further improves the ability of the model to extract feature information by cascading residuals. The proposed EMSA module allows us to apply Transformer to aggregate global contextual information from multiple scales in the encoder and decoder. The experimental results show that the HUNet model achieves the best performance on all considered evaluation metrics. In qualitative comparisons, this is illustrated in various cases in which HUNet accurate segmentation difficult samples (e.g., Fig.10 and Fig.11).

Despite obtaining promising results, our research has some limitations. Specifically, in some cases with several liver cysts in close proximity of the kidney, the automatic segmentation method may overestimate the kidney volume due to the inclusion of liver cysts within the segmented kidney region. To potentially overcome this issue, the proposed method can be trained on 3D volumes of polycystic kidneys. Another limitation of our study is that we only analyzed MRI images. Future work is needed to extend the proposed method to CT, by training the CNN and specifically tune the parameters used during training for CT images. As future work, the automated method can be trained on other affected organs.

VI. CONCLUSION

This paper proposes an automatic segmentation model of HUNet based on multi-module hybrid for ADPKD segmentation of MRI images. It is characterized in that a wavelet pooling module is designed, which introduces the multi-scale analysis of wavelet decomposition into the CNN framework, which reduces the loss of feature information during the pooling process of the network. The cascade residual module is used to further multiplex the feature information between layers, which improves the model's ability to extract features. The novel self-attention allows us to extend operations at different levels of the network in both encoder and decoder

for better capturing long-range dependencies. Validated on the test set, the HUNet model Dice, Jaccard, Precision and Recall are 0.915, 0.844, 0.918 and 0.913 respectively, thus demonstrating that the model can effectively segment ADPKD and its evaluation metrics outperform other CNN models with similar structures. The total kidney volumes derived using manual and HUNet segmentation methods were in high agreement. The percentage difference in total kidney volume values measured using the manual and HUNet methods was only 0.4%.

These findings demonstrate an automated segmentation method that measures TKV as accurately as manual tracing. The method may facilitate those studies in which TKV measurements are needed to assess disease severity, disease progression, and treatment response.

DATA AVAILABILITY

The data source of this paper is provided by the partner Shanghai Changzheng Hospital, mainly for the kidney segmentation task of ADPKD. The datasets used in this study are not publicly available due to specific requirements to regulate privacy protection.

The study was conducted in accordance with Ethics Committee of Shanghai Changzheng Hospital (CZEC(2013)-12).

REFERENCES

[1] A. B. Chapman, J. E. Bost, V. E. Torres, L. Guay-Woodford, K. T. Bae, D. Landsittel, J. Li, B. F. King, D. Martin, L. H. Wetzel, M. E. Lockhart, P. C. Harris, M. Moxey-Mims, M. Flessner, W. M. Bennett, and J. J. Grantham, "Kidney volume and functional outcomes in autosomal dominant polycystic kidney disease," *Clin. J. Amer. Soc. Nephrol.*, vol. 7, no. 3, pp. 479–486, Mar. 2012.

[2] G. M. Fick-Brosnahan, M. M. Belz, K. K. McFann, A. M. Johnson, and R. W. Schrier, "Relationship between renal volume growth and renal function in autosomal dominant polycystic kidney disease: A longitudinal study," *Amer. J. Kidney Diseases*, vol. 39, no. 6, pp. 1127–1134, Jun. 2002.

[3] A. B. Chapman and W. Wei, "Imaging approaches to patients with polycystic kidney disease," *Semin. Nephrol.*, vol. 31, no. 3, pp. 237–244, May 2011.

[4] F. Flinter, "Autosomal dominant polycystic kidney disease," *J. Med. Genet.*, vol. 369, no. 7, pp. 1287–1301, 1996.

[5] W. Thong, S. Kadoury, N. Piché, and C. J. Pal, "Convolutional networks for kidney segmentation in contrast-enhanced CT scans," *Comput. Methods Biomech. Biomed. Eng., Imag., Visualizat.*, vol. 6, no. 3, pp. 277–282, May 2018.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[7] R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," *Proc. SPIE*, vol. 4322, pp. 1337–1346, Jul. 2001.

[8] V. Daum, H. Helbig, R. Janka, K.-U. Eckardt, and R. Zeltner, "Quantitative measurement of kidney and cyst sizes in patients with autosomal dominant polycystic kidney disease (ADPKD)," in *3rd Russian-Bavarian Conf. Biomed. Eng.*, vol. 1, J. Hornegger, et al., Ed., 3rd ed. Erlangen, Germnay, 2007, pp. 111–115.

[9] D. Racimoraa, P.-H. Vivierb, H. Chandaranaa, and H. Rusinek, "Segmentation of polycystic kidneys from MR images," *Proc. SPIE*, vol. 7624, Mar. 2010, Art. no. 76241W.

[10] K. Sharma, L. Peter, C. Rupprecht, A. Caroli, L. Wang, A. Remuzzi, M. Baust, and N. Navab, "Semi-automatic segmentation of autosomal dominant polycystic kidneys using random forests," 2015, *arXiv:1510.06915*.

[11] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, Dec. 2002.

[12] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Tech. Rep., 1998.

[13] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*.

[16] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical transformer: Gated axial-attention for medical image segmentation," 2021, *arXiv:2102.10662*.

[17] Y. Zhang, H. Liu, and Q. Hu, "TransFuse: Fusing transformers and CNNs for medical image segmentation," 2021, *arXiv:2102.08005*.

[18] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. S. Torr, and L. Zhang, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," 2021, *arXiv:2012.15840*.

[19] A. Goel, G. Shih, S. Riyahi, S. Jeph, H. Dev, R. Hu, D. Romano, K. Teichman, J. D. Blumenfeld, I. Barash, I. Chicos, H. Rennert, and M. R. Prince, "Deployed deep learning kidney segmentation for polycystic kidney disease MRI," *Radiol., Artif. Intell.*, vol. 4, no. 2, Mar. 2022, Art. no. e210205.

[20] A. Raj, F. Tollens, L. Hansen, A.-K. Golla, L. R. Schad, D. Nörenberg, and F. G. Zöllner, "Deep learning-based total kidney volume segmentation in autosomal dominant polycystic kidney disease using attention, cosine loss, and sharpness aware minimization," *Diagnostics*, vol. 12, no. 5, p. 1159, May 2022.

[21] Y. Kim, C. Tao, H. Kim, G.-Y. Oh, J. Ko, and K. T. Bae, "A deep learning approach for automated segmentation of kidneys and exophytic cysts in individuals with autosomal dominant polycystic kidney disease," *J. Amer. Soc. Nephrol.*, vol. 33, no. 8, pp. 1581–1589, Aug. 2022.

[22] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[23] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Proc. Int. Conf. Rough Sets Knowl. Technol.*, 2014, pp. 364–375.

[24] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," in *Proc. ICLR*, 2013, pp. 1–9.

[25] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 710–732, Jul. 1992.

[26] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2091–2110, Dec. 1989.

[27] Y. Tian, H. Cui, Z. Pan, J. Liu, S. Yang, L. Liu, W. Wang, and L. Li, "Improved three-dimensional reconstruction algorithm from a multifocus microscopic image sequence based on a nonsubsampled wavelet transform," *Appl. Opt.*, vol. 57, no. 14, pp. 3864–3872, 2018.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[29] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. IEEE 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[30] D. Karimi and S. E. Salcudean, "Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 499–513, Feb. 2020.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[33] A. Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., 2012.

[34] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Oct. 2018, pp. 327–331.

HAOYANG CUI received the B.S. degree in mechanical engineering from the Henan University of Science and Technology, Henan, China, in 2016, and the M.S. degree in mechanical engineering from Shanghai University, Shanghai, China, in 2019. He received the title of Intermediate Engineer, in 2021. He is currently an Algorithm Researcher with Shanghai Aitrox Technology Company Ltd. His research interests include deep learning, machine learning, and the application of artificial intelligence in the field of medical imaging. He has published more than ten articles in related fields.

YIYI MA received the M.D. degree in nephrology from Second Military Medical University. He is currently an Associate Professor with the Nephrology Department, Shanghai Changzheng Hospital, Shanghai, China. He is also the Vice Director of the Nephrology Department, Kidney Institute, Changzheng Hospital. He has published 12 SCI/SCIE index papers so far. His research interests include hereditary kidney disease, chronic kidney disease management, and nephrolithiasis medical diagnosis and treatment.

MING YANG received the M.D. degree from Second Military Medical University, Shanghai, China. He is currently an Attending Physician with Shanghai Changzheng Hospital, Shanghai. He is also with the Kidney Institute, Department of Nephrology, Shanghai Changzheng Hospital. He has published more than ten SCI/SCIE indexed papers so far. His research interests include the pathogenesis of autosomal dominant polycystic kidney disease (ADPKD) and drug treatment, pathogenesis, and therapeutic targets of acute kidney injury.

YANG LU received the B.Sc. degree in life science from The Hong Kong University of Science and Technology, in 2014, and the M.Sc. degree in biology from the University of Munich, Germany, in 2016. He is currently a Scientist with Shanghai Aitrox Technology Company Ltd. He is also focusing on the application of artificial intelligence in medical imaging.

MINGZI ZHANG received the Ph.D. degree in engineering from Tohoku University, Japan, and the Ph.D. degree in biomedical sciences from Macquarie University, Australia. He is currently a Research Scientist with Shanghai Aitrox Technology Company Ltd., Shanghai, China. He has published more than 20 articles in and served as the external peer-reviewer for various esteemed academic journals. His research interests include computational modeling of the cardiovascular system and diseases, optimization and virtual deployment of endovascular devices, and surgical implants.

LILI FU received the degree in microbiology from the Department of Biology, Northwest University, Shanxi, China. She is currently a Technician with the Division of Nephrology, Kidney Institution, Shanghai Changzheng Hospital.

CHICHENG FU received the Ph.D. degree from the University of California at Berkeley, USA. He is currently the Chief Scientist of Shanghai Aitrox Technology Company Ltd., Shanghai, China. He has published more than 20 SCI papers (with total impact factors more than 240 and citations more than 3000), including publications in *Science*, *PNAS*, and *Advanced Materials* as the first or co-first author. His research interests include AI in medicine, nano-fluidics, single stem cell, and gene regulation.

BEILIN SU received the degree from Second Military Medical University, Shanghai, China. Her major is pharmacology. She is currently a Technician with the Division of Nephrology, Kidney Institution, Shanghai Changzheng Hospital. She has published more than three SCI/SCIE indexed papers so far.

CHUAN HE received the B.S. degree in international economics and trade from the Antai College of Economics and Management, SJTU, Shanghai, China, in 2010. She is currently a Global Partner with Fosun Group and the Chairperson of Shanghai Aitrox Technology Company Ltd.

CHENG XUE received the Medical Doctor degree from Second Military Medical University, Shanghai, China. He is currently an Associate Professor with the Department of Nephrology, Shanghai Changzheng Hospital. He has published more than 50 SCI/SCIE indexed papers so far. He is good at the diagnosis and treatment of glomerulonephritis, chronic renal failure and its complications, and hereditary nephropathy.

CHANGLIN MEI is currently a Professor, a Ph.D. Supervisor, and the Director of the Nephrology Institute, Shanghai Changzheng Hospital, Shanghai, China. He is also the Director of the Shanghai Nephrology Clinical Quality Control Center, the Vice President of the Nephrology Physician Branch of the Chinese Medical Doctor Association, and the Vice President of the Rare Disease Branch of the China Research Hospital Association. He undertook 23 research topics, including key and general projects of the National Natural Science Foundation of China, major national science and technology projects, and major research projects of the Shanghai Municipal Science and Technology Commission. He is the Editor-in-Chief and the Associate Editor-in-Chief of 23 monographs, published 427 papers, and published more than 100 SCI papers in *NEJM*, *JASN*, *KI*, *NDT*, and other magazines. He obtained one national new drug certificate and five invention patents in China and the USA. He has won 14 major science and technology and medical achievement awards, including the Second Prize of the National Science and Technology Progress Award, the First Prize of the National Ministry of Education Science and Technology Progress Award, the First Prize of the Military Medical Achievement Award, and the First Prize of the Shanghai Science and Technology Progress Award.

SHUWEI SONG received the M.D. degree from Second Military Medical University, Shanghai, China. He is currently with the Kidney Institute, Department of Nephrology, Shanghai Changzheng Hospital, Shanghai. He is executively in charge of the Kidney Institute, Shanghai Changzheng Hospital. He undertook and participated in nine research topics, including the National Natural Science Foundation of China and the Foundation of Shanghai Science and Technology Commission. He has published more than 15 SCI/SCIE indexed papers so far. He obtained more than ten invention patents in China. His research interests include the drug treatment and pathogenesis of autosomal dominant polycystic kidney disease (ADPKD) and diagnosis of acute kidney injury.