Computer Methods and Programs in Biomedicine 221 (2022) 106861

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

A deep learning-based precision volume calculation approach for kidney and tumor segmentation on computed tomography images☆

Chiu-Han Hsiao[a], Tzu-Lung Sun[b], Ping-Cherng Lin[a], Tsung-Yu Peng[b], Yu-Hsin Chen[b], Chieh-Yun Cheng[b], Feng-Jung Yang[c,*], Shao-Yu Yang[d], Chih-Horng Wu[e], Frank Yeong-Sung Lin[b], Yennun Huang[a]

[a] Research Center for Information Technology Innovation, Academia Sinica, Taipei City, Taiwan, ROC
[b] Department of Information Management, National Taiwan University, Taipei City, Taiwan, ROC
[c] Department of Internal Medicine, National Taiwan University Hospital Yunlin Branch, Douliu City, Yunlin County; School of Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan, ROC
[d] Department of Internal Medicine, National Taiwan University Hospital, Taipei City, Taiwan, ROC
[e] Department of Medical Imaging, National Taiwan University Hospital, Taipei City, Taiwan, ROC

ARTICLE INFO

Article history:
Received 28 September 2021
Revised 24 March 2022
Accepted 7 May 2022

Keywords:
Deep learning
Kidney volume
Preprocessing
Semantic segmentation

ABSTRACT

Previously, doctors interpreted computed tomography (CT) images based on their experience in diagnosing kidney diseases. However, with the rapid increase in CT images, such interpretations were required considerable time and effort, producing inconsistent results. Several novel neural network models were proposed to automatically identify kidney or tumor areas in CT images for solving this problem. In most of these models, only the neural network structure was modified to improve accuracy. However, data pre-processing was also a crucial step in improving the results. This study systematically discussed the necessary pre-processing methods before processing medical images in a neural network model. The experimental results were shown that the proposed pre-processing methods or models significantly improve the accuracy rate compared with the case without data pre-processing. Specifically, the dice score was improved from 0.9436 to 0.9648 for kidney segmentation and 0.7294 for all types of tumor detections. The performance was suitable for clinical applications with lower computational resources based on the proposed medical image processing methods and deep learning models. The cost efficiency and effectiveness were also achieved for automatic kidney volume calculation and tumor detection accurately.

1. Introduction

Autosomal dominant polycystic kidney disease, also known as polycystic kidney syndrome, is a genetic disease in which the renal tubules become structurally abnormal, resulting in the outgrowth of multiple cysts and a decline in renal function. The outgrowth of cysts causes the kidneys to swell and change shape. Therefore, the measurement of total kidney volume (TKV) is adopted as a biomarker for the disease status evaluation [1–3]. High-resolution computed tomography (CT) images can provide the necessary anatomical details to diagnose the disease progression, such as Alzheimer's disease or mild cognitive impairment (MCI) in neural Computing and applications [4]. However, interpreting these images requires a great deal of time and effort to identify the locations of cysts manually. Therefore, deep learning algorithms [5] for semantic segmentation are being developed to simultaneously locate and delineate kidney areas [6]. These methods enable automatic and precise segmentation of kidney and tumor, thus facilitating pathological diagnosis and therapy. Further, they can increase labor and time efficiency. Finally, these methods provide consistent and highly accurate identification results.

Kidney feature extraction usually relies on sharp intensity variations among voxels around the kidney boundary in CT images. However, some phenomena such as low contrast, noise, opacity, and anisotropy may arise when capturing CT images. These deteriorate the segmentation results when ambiguous voxels appear near the boundaries of organs. The extracted features cannot fully

☆ This paper is a part of the results of the research project funded by National Taiwan University Hospital Yunlin Branch.
* Corresponding author.
E-mail address: fongrong@ntu.edu.tw (F.-J. Yang).

https://doi.org/10.1016/j.cmpb.2022.106861

indicate the outline of kidneys. Further, the tininess of tumors and the resemblance in texture between tumors and kidneys can complicate the identification process. These problems can affect the TKV measurement outcomes. As a result, practical operations remain challenging [7,8].

Semantic segmentation algorithms derived from convolutional neural networks (CNNs) were developed to solve these problems. Modified CNNs can overcome the shortage of training samples in medical image datasets to perform accurate segmentation without any manual intervention [6,9].

To obtain better-trained models, data pre-processing plays a crucial role in the entire segmentation process [10]. Cleaned and enhanced data, including the training dataset and ground truth, can improve segmentation performance in specific medical applications [4]. In this paper, three pre-processing procedures were analyzed and adapted to improve the accuracy rate of kidney segmentation systematically. The first step uses statistical Hounsfield unit (HU) windowing to adjust the grayscale to remove noise and suppress adjacent organs' intensities. The second step discriminates whether kidney areas exist in CT image slices. This method is used to evaluate the influence related to class imbalance. The third step analyzes the difference between single and multiple image labels. Two modified U-Nets with different encoder models were used to test these pre-processing data. The preliminary results demonstrated that these pre-processing methods could improve the accuracy rate effectively. The dice score was increased from 0.9436 to 0.9524. Further, the TKV calculation was evaluated. The proposed methods enabled automatic and precise segmentation of kidneys and tumors to provide consistent and highly accurate identification results in clinical applications.

The remainder of this paper is organized as follows: Section 2 reviews related works that identified the pre-processing methods and deep learning models. Then, Section 3 introduces the proposed methods, models, and pre-processing procedures to determine a kidney and tumor segmentation solution. Section 4 presents experiments to evaluate the performance metrics and validate the pre-processing procedures. Finally, discussions and conclusions are drawn in Sections 5 and 6.

2. Related work

Recently, CNNs have shown promise for classification, segmentation, localization, and detection in various medical images, thereby supporting clinicians in disease diagnosis [4]. The better the quality of the training data fed into the model, the better the obtained model is [11,12]. Depending on the characteristics of distinct medical images, proper pre-processing methods such as denoising and contrast enhancement are selected to eliminate noise and suppress interference [13]. Another preliminary step is intensity normalization, in which the intensity distribution is transformed or filtered and set in a specific range [4,11,14] to improve the optimization [6,11]. Based on the previous studies, the pre-processing methods and deep learning models on the leader board of KiTS19 Challenge are also surveyed for the object detection on medical images accurately in clinical applications [15].

2.1. Pre-processing digital

Methods CT images were obtained by normalizing the X-ray attenuation coefficients, depending on the attenuation in water and air. The scale is expressed in HU with usually ranges from -1024 to 3071. The appropriate HU range can be selected to map 256 gray values for displaying specific organs; it is a fast and straightforward way to perform linear contrast enhancement using a linear transfer function. The technique is called HU windowing [16]. Many studies have used HU windowing to make target organs prominent and remove the irrelevant background by observation. HU values have been set in distinct ranges for the same target, namely $[-100, 400]$ in Christ et al. [17], $[-75, 175]$ in Zhang et al. [18], and $[-200, 250]$ in Chen et al. [19]. The current study presents a statistical approach to enhancing the volumes of interest by stretching gray values to clean the training datasets for the CNN model. According to the quality levels of various datasets, nonlinear histogram-based and filtering techniques are often employed to eliminate noise to acquire high-quality images [20,21].

CT images were captured in the slice model. When the abdomen is scanned, the sliced images contain the entire abdominal cavity. However, the kidneys are only small parts of the entire image cube. Most portions of the sliced images do not include the kidney areas; therefore, the imbalanced class issues should be verified [22]. Cruz et al. applied the AlexNet model for separating kidney and non-kidney slices. A recovery method was used to correct the discontinuity among slices [6]. The kidney-containing slices selected as training inputs and the influence of the removed background voxels are assessed in our study. In addition, the ground truth datasets could affect the model performance. In the KiTS Challenge, background voxels of the ground truth datasets were labeled as 0; kidney voxels were labeled as 1, and kidney tumor voxels were labeled as 2. Based on kidney tumors have small and irregular shapes, the accurate identification of the kidney volume depends on whether kidney tumors are considered parts of the entire kidney area during model training. Isensee et al. [23] treated the kidney and tumor labels as foreground voxels and the others as background voxels. The kidney dice score reached 0.9737, and the tumor dice score reached 0.8573; their work won first place in the challenge. Further, Hou et al. [24] used a cascaded method that segmented the kidney areas in advance. Then, the tumors were considered foreground voxels, and the extracted kidney areas were considered background voxels. The kidney dice score reached 0.9674, and the tumor dice score reached 0.8454; this work won second place in the challenge. Overall, the labeling method seemed to significantly influence the accuracy of the segmentation results obtained in these studies.

2.2. Deep learning models

The U-Net model [25] is a popular architecture for the semantic segmentation of medical images, and many U-Net variants have been developed to improve dice scores. In the KiTS19 Challenge, Isensee et al. [23] and Hou et al. [24] both used a three-dimensional (3D) U-Net architecture. Specifically, Isensee et al. [23] adopted a residual 3D U-Net, and Hou et al. [24] used multistage processes to obtain precise kidney locations and improve tumor segmentation results by using a generic 3D U-Net. Mu et al. [26] used a multistage method and a residual 3D U-Net called VB-Net, which applied a V-Net architecture [27] but replaced the conventional convolutional layers with a bottleneck structure. This model won third place in the competition. Xi et al. [28] proposed Cascade U-ResNets to simultaneously perform liver and lesion segmentation. Cascaded methods are often used to shrink the target area and increase the accuracy of tumor segmentation results [29]. The emerging encoder-decoder architecture downsamples the image resolution for the encoder to extract details. Then, the decoder upsamples the feature map to recover spatial information [30]. Hong et al. [31] used an encoder-decoder model derived from U-Net; EfficientNet substituted the encoder to obtain a better performance than that of the original U-Net model. Because of the limitation of computing resources, some studies have designed newer and more flexible architectures [32,33]; and others have proposed the use of transfer learning to retain pre-trained weights from similar tasks [3,31,34]. The goal is to accelerate the computation and overcome hardware limitations. For achieving higher dice scores,

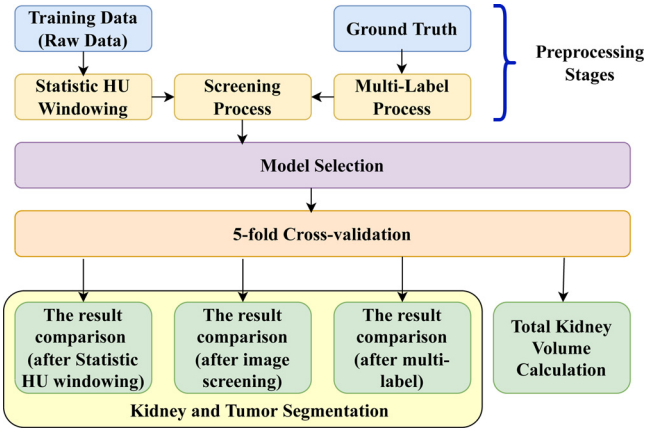

Fig. 1. The proposed pre-process and verification flow diagram.

some studies have adopted complicated 3D models and deep reinforcement learning (DRL) models to increase performance, respectively [12,35].

2.3. Summary

This paper systematically proposes pre-processing procedures, including HU windowing, image screening, and labeling methods for kidney and tumor segmentation, regarding the previous studies mentioned above. The deep learning model is selected by U-Net architecture, ResNet-41 or EfficientNet as an encoder modified by [36,37] and Feature Pyramid Network (FPN) as a decoder [38]. A series of optimization steps are evaluated and developed, including windowing selection, data screening, and labeling selections to obtain the optimal hyperparameters. Furthermore, the encoder and decoder combinations for the deep learning design yield the best performance. It is validated by five-fold cross-validation processes to evaluate the accuracy of kidney and tumor detection.

3. Proposed methods

The quality of the training datasets significantly affects the model performance. When proper pre-processing steps are conducted, the models show a noticeable improvement. Fig. 1 shows the proposed pre-processing and verification procedures. In pre-processing, processed training data are trained using the ResNet model for analyzing performance metrics. The procedures can be divided into statistical HU windowing, image screening, and alternative process labeling. Then, the ResNet and EfficientNet models evaluate the pre-processing performance with varying pre-trained parameters as initial weights.

3.1. Statistical HU windowing

According to their locations, the number of kidney or tumor voxels differs in each CT image slice. The proposed windowing method counted the HU values of all kidney voxels within the entire training dataset in terms of the regions enclosed by the ground truth dataset to sharpen the kidney areas or tumor areas, so on so forth. The distribution of the HU values was similar to the normal distribution, as shown in Fig. 2. Therefore, the window level can be obtained by calculating the middle number of the HU values of all kidney or tumor voxels. The window width was the range of HU values for display. The voxels showed different gray levels within the range of $HU_{\min}$ to $HU_{\max}$. The voxels appeared white and black when the HU values were larger than $HU_{\max}$ and smaller

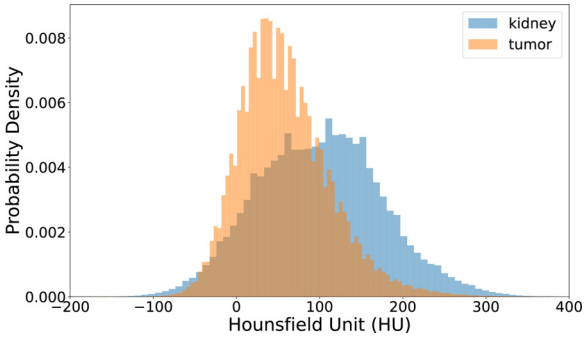

Fig. 2. The distribution curve of the HU values of all kidney or tumor voxels.

than $HU_{\min}$, respectively. The adjustment formulas are as follows:

$$HU_{\max} = \text{Window Level} + \frac{\text{Window Width}}{2} \tag{1}$$

$$HU_{\min} = \text{Window Level} - \frac{\text{Window Width}}{2} \tag{2}$$

The window width and window level can be derived from (1) and (2) as

$$\text{Window Level} = \frac{HU_{\max} + HU_{\min}}{2} \tag{3}$$

$$\text{Window Width} = HU_{\max} - HU_{\min} \tag{4}$$

The standard deviation (SD) and mean of the HU values of all kidney or tumor voxels can be determined using $HU_{\min}$ and $HU_{\max}$. Our observations suggested that the HU values in the range of window level ($\pm$ 3$\times$standard deviation) can represent most kidney or tumor areas and remove outliers. In our study, various standard deviations were selected to set distinct windowing ranges. We designed three situations with one, two, and three standard deviations to obtain the training data. The processed data were fed into the ResNet model to evaluate the effects.

3.2. Image screening

The CT image cube showed that the number of slices with kidney or tumor voxels was less than that of non-kidney areas. In the KiTS19 Challenge, kidney-containing slices accounted for only 36% of the dataset, and non-kidney areas accounted for 64% of the dataset. Tumor-containing slices accounted for only 13% of the dataset. However, the majority of the CT image slices do not show kidney or tumor voxels, irrespective of whether class imbalance affects the model performance. For verifying the situations mentioned above, the image-screening procedure was applied to reduce the dimensions of the CT images. We planned three scenarios for evaluating the effects: (1) Retaining all slices and adopting a whole CT image cube as input data, (2) filtering out images without kidneys or tumors, and (3) selecting slices with at least 1% kidney voxels or 0.01% tumor voxels.

3.3. Labeling alternative

Based on tumors are small, the semantic segmentation model concerning single or multiple targets may affect identification accuracy. Labeling the ground truth dataset in the KiTS19 Challenge obeyed the following rules: non-kidney, kidney, and tumor areas were labeled as 0, 1, and 2, respectively. Fig. 3 shows different labeling methods. Fig. 3(A) shows the original kidney image slice. Fig. 3(B) shows the kidney areas and tumors by using distinct labels. Because tumors are mostly located at the edges of the kidney or are enclosed in kidney areas, removal of kidney tumors

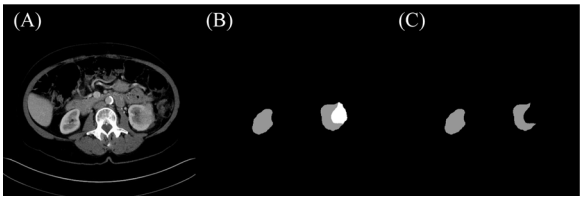

Fig. 3. (A) The original CT image slice (B) The bright area denotes the kidney tumor. The gray parts represent kidneys without tumors. The background is black. (C) The kidney areas without the tumor show an irregular shape.

gives kidneys a concave form, as shown in Fig. 3(C). The proposed method assessed the influence of the label alternative. The labels of the kidney areas and tumors were merged into one label such that, for example, the label {0,1,1} indicated that the kidney and tumor are both labeled as 1.

3.4. Model selection

U-Net is commonly applied in medical image segmentation. An encoder-decoder architecture is a characteristic of various applications. During training stages, transfer learning accelerated the computation. Based on the literature review, the difference in the U-Net encoder affects the performance and efficiency of the model. The ResNet-41 model uses pre-trained weights from the ImageNet dataset, whereas the EfficientNet model initializes pre-trained weights acquired using a noisy-student dataset. This study selected the ResNet-41 and EfficientNet as the encoder for comparison and evaluation with and without pre-processed datasets.

3.4.1. ResNet-41

The ResNet-41 is a variant of the ResNet model [36] used in this work. Owing to limited computational power, ResNet-41 has a fair number of layers compared with other models and can retain feature maps to some extent. The architecture of the proposed U-Net with ResNet-41 as an encoder is shown in Fig. 4.

3.4.2. EfficientNet

Our study also introduced the EfficientNet model [37] as the encoder with the newest parameters obtained by training a noisy-student dataset for performance evaluation. With EfficientNet-B0 used as a baseline model, a series of models (B1 to B7) was developed. The EfficientNet model of the noisy-student version was based on the EfficientNet model of the ImageNet version [39]. Through several iterations of knowledge distillation, the noisy-student version of EfficientNet can finally be constructed.

3.4.3. Feature pyramid network

The feature pyramid network (FPN) is conducted as a decoder in this paper. FPN generates prediction layers at multiple scales. The FPN comprises a bottom-up pathway, a top-down pathway, and lateral connections [38]. Based on the characteristics of kidney tumors of various types and have irregular shapes, replacing the decoder from U-Net to FPN can improve the segmentation performance.

4. Experiments

4.1. Datasets

KiTS19 provides the CT images of 210 patients with a resolution of 512–512. Each CT image has three annotations: 0 for background, 1 for kidney, and 2 for kidney tumor. The kidney segmentation dataset used in this study was obtained from KiTS19

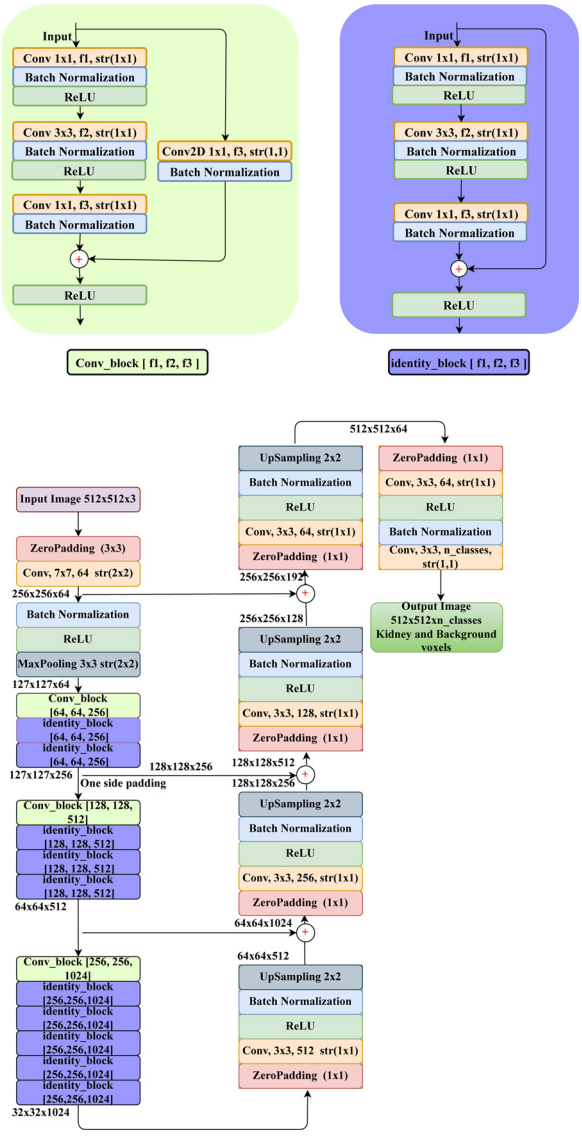

Fig. 4. The architecture of proposed U-Net with ResNet-41 as an encoder.

[15] (https://kits19.grand-challenge.org/); the challenge aims to determine the best solution for the semantic segmentation of renal kidney tumors in CT images. The dataset contained 210 training cases and 90 testing cases. Because only the 210 training cases included annotation data, the segmentation data of the 90 testing cases were private and could not be accessed. Therefore, the 210 training cases were split into training and testing data in this study because complete annotation information was available for these cases. In these training cases, the number of slices ranged from 29 to 1059, slice thickness ranged from 0.5 to 5 mm, pixel width ranged from 0.437 to 1.041 mm, TKV (including tumors and two sides of each kidney) ranged from 192 to 1962 mL, and tumor volume ranged from 0.82 to 1464 mL. For evaluating renal anatomical characteristics, cases 165, 194, 213, 293, and 294 were solitary (i.e., with only one kidney). Moreover, cases 5 and 151 featured a horseshoe kidney (left and right kidneys joined just above the spine). Labels 0, 1, and 2 indicated the background in the labeled data, specific kidney regions, and kidney tumor regions, respectively. The data format was NIfTI. The image order was channel first, and the voxel spacing was recorded in each case's header file: kidney tumor type, contrast medium injection duration, and scanner variable case by case. The data diversity of the

Table 1
Distinct HU_{min}, HU_{max}, window widths and window levels correspond to a specific HU value distribution range of the kidney areas in the training set of KiTS19 Challenge data set.

Range	HU_{min}	HU_{max}	Window width	Window level
mean ± 3 SD	−130	330	460	100
mean ±2 SD	−53	254	306	100
mean ± 1 SD	24	177	153	100

Table 2
The kidney dice scores of five-fold cross-validation with various HU ranges .

Kidney Dice Score	One SD	Two SD	Three SD
1st fold	0.9307	0.9404	**0.9455**
2nd fold	0.9451	0.9573	**0.9618**
3rd fold	0.9337	0.9428	**0.9495**
4th fold	0.9444	**0.9607**	0.9526
5th fold	0.9433	**0.9548**	0.9520
Average	0.9394	0.9512	**0.9523**

KiTS19 dataset made performing tumor and kidney segmentation challenging.

4.2. Performance evaluation

The dice score is set as the performance metric to evaluate segmentation performance. The dice score indicates how the predicted mask overlaps with the accurate mask. The equation is shown as (5). A true positive (TP) is the number of valid positive pixels of the prediction class. A false positive (FP) incorrectly predicts the positive class. However, a false negative (FN) mispredicts the negative class. Therefore, a higher dice score indicates a more precise segmentation result. This study used five-fold cross-validation to evaluate the results and obtained the validation dice score for each case. Because the validation dice scores did not follow a normal distribution, we applied the Wilcoxon signed-rank test, a nonparametric test method, to evaluate whether the experimental results were significant [40].

$$\text{Dice} = \frac{2 \times \text{TP}}{(\text{TP} + \text{FP}) + (\text{TP} + \text{FN})} \tag{5}$$

4.3. Training and analysis

4.3.1. Statistical HU windowing

Various HU ranges using the formulas given in Section 3.1 can be calculated with different standard deviations. Thus, the window width and window level can be derived. Table 1 shows the range of HU, window width, and window level. The proposed windowing technique applied different window widths and window levels to the KiTS19 Challenge dataset. According to Table 1, the window widths and levels (width, level) were (460, 100), (306, 100), and (153, 100). This study used the U-Net model with the ResNet-41 model as an encoder to evaluate the three processed datasets. Table 2 shows the results of different windowing methods for five-fold cross-validation.

Table 2 shows that the dataset with windowing in the range of three standard deviations had the highest average dice score, followed by the datasets with windowing in the ranges of two standard deviations and one standard deviation. Moreover, the statistical test results in Table 3 indicated that the datasets with windowing in the ranges of two and three standard deviations significantly outperformed those with windowing in the range of one standard deviation. Because the dataset with windowing in the range of three standard deviations had the highest average dice score, we used it to perform the subsequent experiments. Figs. 5 and 13 show the significant comparisons with various standard deviations for kidney and tumor segmentation, respectively.

Table 3
The statistic test with various standard deviations as HU range .

p-value	One SD	Two SD	Three SD
One SD	–	**0.000***	**0.000***
Two SD	**0.000***	–	1.000
Three SD	**0.000***	1.000	–

Table 4
The kidney dice scores of five-fold cross-validation with various image screening methods .

Kidney dice score	All slices	Slices without kidneys	Slices with kidney voxels ≥ 1%
1st fold	**0.9455**	0.9122	0.8317
2nd fold	**0.9618**	0.9305	0.8879
3rd fold	**0.9495**	0.9238	0.8604
4th fold	**0.9526**	0.9484	0.9119
5th fold	**0.9520**	0.9047	0.8726
Average	**0.9523**	0.9240	0.8729

Table 5
The statistic test with various image screening methods .

p-value	All slices	Slices without kidneys	Slices with kidney voxels ≥ 1%
All	–	**0.000***	**0.000***
Black removed	**0.000***	–	**0.000***
≥ 1%	**0.000***	**0.000***	–

Table 6
The precision results of five-fold cross-validation with various image screening methods .

Precision	All slices	Slices without kidneys	Slices with kidney voxels ≥ 1%
1st fold	**0.9650**	0.8944	0.7558
2nd fold	**0.9701**	0.9004	0.8354
3rd fold	**0.9771**	0.9032	0.7934
4th fold	**0.9722**	0.9440	0.8796
5th fold	**0.9626**	0.8795	0.8163
Average	**0.9708**	0.9036	0.8156

4.3.2. Image screening

As noted in the previous section, we windowed the CT image within the range of three standard deviations. Based on this dataset, we further filtered out the CT images with small regions of interest. As mentioned in Section 3.2, the three filter methods are listed as follows: (1) Retaining all slices, (2) filtering out images without kidneys, and (3) filtering out images with kidney voxels accounting for less than 1% of the area or tumor voxels accounting for less than 0.01%. We continued to use U-Net with ResNet-41 as the segmentation model. Table 4 and Fig. 13 shows the results.

According to Table 4, using the dataset with all CT scans provided the best performance. The statistical results in Table 5 show that the performances differed significantly. Fig. 6 shows the significant comparisons with various image-screening methods. The model used all CT scans for training yielded the highest dice score. We speculated that if we used only images with kidneys to train the model, the model might misclassify non-kidney areas as kidney areas. To testify our conjecture, we further evaluated the precision and recall of the segmentation results.

From Tables 6 and 7, as the percentage of kidney area in an image increased, the recall improved slightly, but the precision decreased significantly. This result validated our argument that when images with kidneys are used to train the model, the model eas-

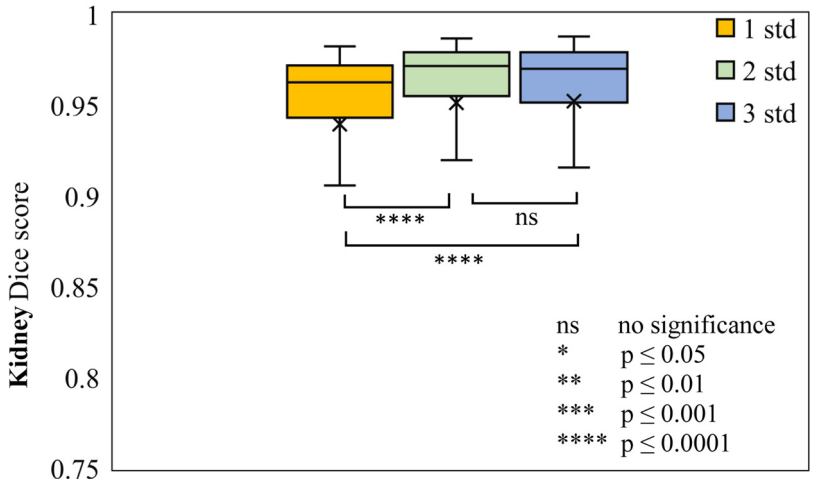

Fig. 5. Significant comparison with various standard deviations. Dice scores are larger than 0.8 shown on the diagram.

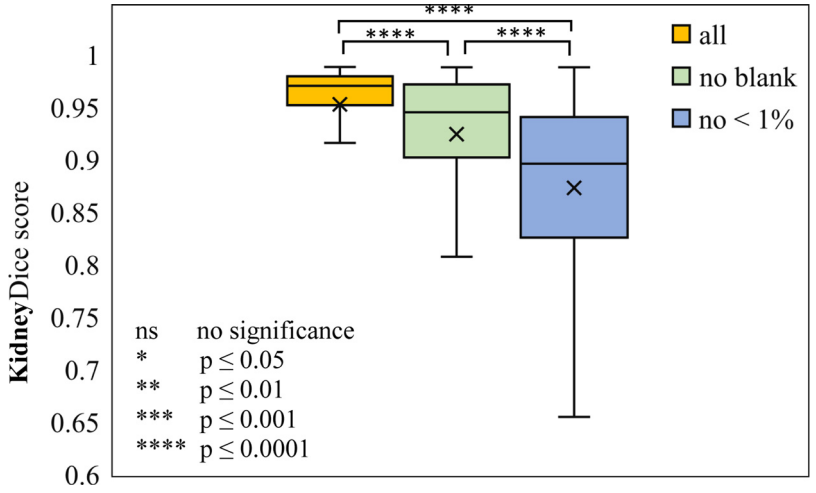

Fig. 6. Significant comparison with various image screening methods.

Table 7
The recall results of five-fold cross-validation with various image screening methods.

Recall	All slices	Slices without kidneys	Slices with kidney voxels ≥ 1%
1st fold	0.9375	0.9487	**0.9587**
2nd fold	0.9571	**0.9668**	0.9609
3rd fold	0.9339	0.9559	**0.9599**
4th fold	0.9404	0.9566	**0.9602**
5th fold	0.9437	0.9458	**0.9485**
Average	0.9342	0.9548	**0.9576**

ily misclassifies non-kidney areas as kidney areas. Although this method can help detect areas that are difficult to identify, it easily misclassifies background areas as kidney areas. In short, it does more harm than good. Fig. 13 is shown the results of tumor segmentation by screening tumor images. It is also revealed that many samples of training images influenced the performance and had higher dice scores than other training configurations.

4.3.3. Labeling alternative

As noted in the previous section, we trained our model with all CT images. Then, we redefined the annotation of the images. In Section 3.3, we discussed whether the kidney tumor area was identified as a kidney area. If it was, the labels of background, kidney, kidney tumor were {0, 1, 1}; otherwise, they were {0, 1, 2}. We used U-Net with ResNet-41 as the segmentation model. The results are shown in Table 8. Table 9 shows that the dataset with the redefined annotation outperformed the dataset with the original annotation. Fig. 7 indicates significant comparisons with label alternatives. The statistical test also revealed a significant difference. Kidney tumors were considered to be a part of the kidney to improve segmentation performance.

Table 8
The kidney dice scores of five-fold cross-validation for label alternative.

Kidney dice score	{0,1,2}	{0,1,1}
1st fold	0.9359	**0.9455**
2nd fold	0.9570	**0.9618**
3rd fold	0.9359	**0.9495**
4th fold	**0.9575**	0.9526
5th fold	0.9450	**0.9520**
Average	0.9463	**0.9523**

Table 9
The statistic test for label alternative.

p-value	{0,1,2}	{0,1,1}
{0,1,2}	–	**0.000***
{0,1,1}	**0.000***	–

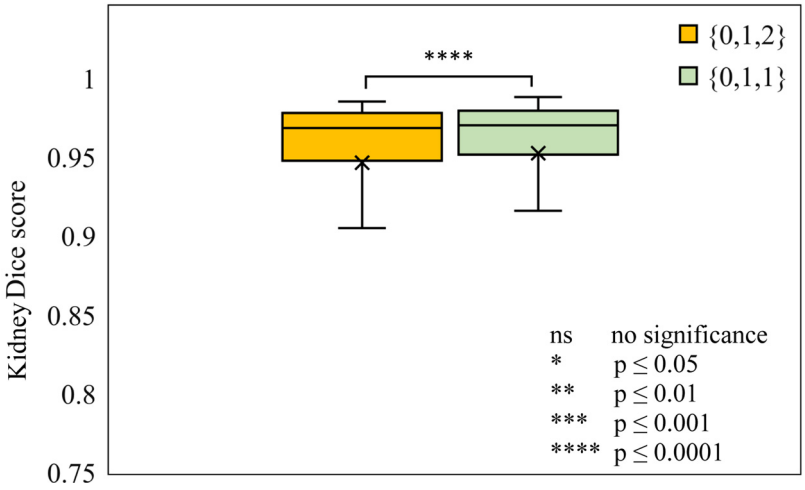

Fig. 7. Significant comparison of varied labeling methods. For clearly showing the difference, data points with dice scores lower than 0.8 do not display on the graph.

Table 10
Evaluation of five-fold cross-validation for kidney segmentation with various encoders on the datasets with pre-processing .

Kidney dice score	ResNet-41	EfficientNet-B4, frozen encoder	EfficientNet-B4, unfrozen encoder	EfficientNet-B7, frozen encoder
1st fold	0.9455	0.9334	**0.9491**	0.9462
2nd fold	0.9618	0.9550	0.9686	**0.9700**
3rd fold	0.9495	0.9543	0.9510	**0.9684**
4th fold	0.9526	0.9614	0.9705	**0.9738**
5th fold	0.9520	0.9583	0.9594	**0.9655**
Average	0.9523	0.9525	0.9597	**0.9648**

Table 11
Evaluation of five-fold cross-validation for kidney segmentation with various encoders on the datasets without pre-processing .

Kidney dice score	ResNet-41	EfficientNet-B4, frozen encoder	EfficientNet-B4, unfrozen encoder	EfficientNet-B7, frozen encoder
1st fold	0.9341	0.4114	**0.9466**	0.3083
2nd fold	0.9363	0.2470	**0.9644**	0.2788
3rd fold	0.9353	0.3211	**0.9379**	0.2403
4th fold	0.9413	0.2193	**0.9650**	0.2992
5th fold	0.9218	0.2784	**0.9541**	0.2327
Average	0.9338	0.2954	**0.9536**	0.2718

4.3.4. Model selection-kidney segmentation

This section follows the experimental results in the previous section; {0, 1, 1} was used to annotate the label image. We applied U-Net with several encoder models, as mentioned in Section 3.4, to compare the segmentation performance of different encoders. Moreover, as discussed in Section 3.4, we conducted transfer learning, which generally achieves the desired performance with a low computational burden due to the pre-trained weights. However, we sought higher performance and expected our model to fit the data best; thus, we also attempted to unfreeze the weights of the pre-trained encoder model during training. In other words, we considered the pre-trained weights to be the initial weights.

As shown in Table 10, EfficientNet performs better than ResNet-41 in most of the folds. Among EfficientNets, EfficientNet-B7 with weights frozen performs the best.

In addition, the experiments are conducted on the datasets with pre-processing and without pre-processing procedures for performance comparison. As shown in Table 10 and Table 11, EfficientNet-B7 with frozen encoder with pre-processing methods performs best. The weights of the encoder can be significantly influenced the performance, which can be found more clearly in Table 11. Without pre-processing methods, the cases of EfficientNet-B4 and EfficientNet-B7 have lower performance than pre-processing ones. The results are revealed that the models can not obtain better features without pre-processing methods. On the other hand, ResNet-41 and EfficientNet-B4 with unfrozen encoder on data pre-processing also performed better than without pre-processing.

Under the statistical test we conducted in Table 12, EfficientNet-B7 with weights frozen has significantly better performance than other models. The diagram of significant comparison for various models can be seen in Fig. 8. Limited by the computing resources of our machine, we did not try EfficientNet-B7 with weights unfrozen. However, the results of EfficientNet-B4 showed that unfreezing the encoder's weights improves the performance; thus, we expected EfficientNet-B7 with weights unfrozen to reach a better performance. Indeed, it is a trade-off between unfreezing the weights to reach better performance and spending much more computing resources.

4.3.5. Model selection-tumor segmentation

This section discusses the possibility of applying semantic segmentation to tumor segmentation. To achieve this goal, EfficientNet-B5 (with decoder FPN) and 3D U-Net were used. How-

Table 12
The statistic test for various models .

p-value	ResNet-41	EfficientNet-B4, frozen encoder	EfficientNet-B4, unfrozen encoder	EfficientNet-B7, frozen encoder
ResNet-41	–	0.128	**0.001***	**0.000***
EfficientNet-B4,frozen encoder	0.128	–	**0.000***	**0.000***
EfficientNet-B4,unfrozen encoder	**0.001***	**0.000***	–	**0.000***
EfficientNet-B7,frozen encoder	**0.000***	**0.000***	**0.000***	–

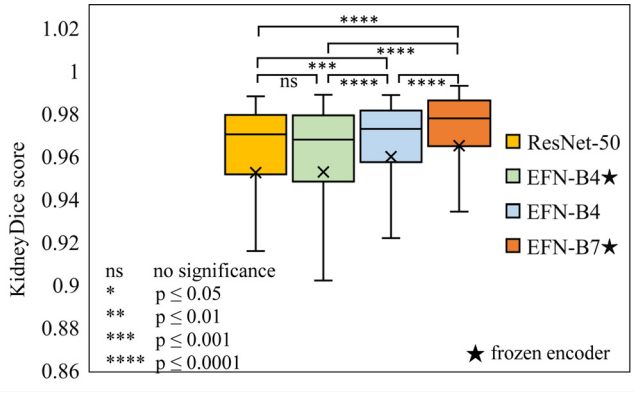

Fig. 8. Significant comparison of various models. For clearly showing the difference, data points with dice scores lower than 0.8 do not display on the graph.

Table 13
Tumor types and the number of cases with annotation in KiTS datasets .

Tumor type	Number of cases
Clear Cell RCC (ccRCC)	143
Papillary	24
Chromophobe	19
Oncocytoma	10
Angiomyolipoma	5
Clear Cell Papillary RCC	4
Mest	2
Spindle Cell Neoplasm	1
RCC unclassified	2
Total	210

ever, owing to the limited memory of the graphics processing unit, the original images had to be processed to a smaller size. The previous kidney segmentation provided a screening function for tumor segmentation. The first stage was to segment the kidney regions and generate smaller images containing the kidney regions. As shown in Fig. 9, the kidney mask was applied to the original image. The second stage was to segment tumor regions out of the kidney regions. Here, 192 × 192 × 192 was an appropriate size that contained the entire kidney region. In these experiments, EfficientNet-B5 and two 3D U-Net sizes were compared. Distinct tumor types and the number of cases are listed in Table 13 with annotation in KiTS datasets. For all tumor categories, clear cell renal cell carcinoma (ccRCC) and other tumor types, the most common tumor types, 3D U-Net (80 × 80 × 80) provided the best dice score. For giant ccRCC tumors, EfficientNet-B5 performed the best. The results also indicate that the models demonstrate better performance when the tumors with a larger size and more cases, as shown in Table 14. Using various models, the datasets with preprocessing procedures show better performance than those without pre-processing procedures, as shown in Fig. 10.

In order to realize the importance of the pre-processing procedures, tumor segmentation results obtained by using the datasets with and without pre-process are shown in Figs. 11 and 12 for the case 25, 26, 29, 31, and 38, respectively. Indeed, data pre-

Table 14
The tumor dice scores for various models on tumor segmentation .

Tumor dice score	EfficientNet-B5	3D U-Net (80-160-160)	3D U-Net (80-80-80)
All	0.5199	0.6760	0.7294
ccRCC	0.5376	0.7101	0.7572
ccRCC ($\leq$ 4 cm)*	0.4214	0.6312	0.7076
ccRCC ($>$ 4 cm)*	0.8616	0.8401	0.8521
Papillary	0.2588	0.3408	0.3556
Chromophobe	0.2066	0.2202	0.2744
Oncocytoma	0.0962	0.0577	0.1204

*Followed by the TNM staging system, 4 cm is the threshold between T1a and T1b [41,42].

processing methods enhance the performance of neural network models. This study depicts the necessary processes or methods before the medical images are applied to the neural network models. The processes are summarized as followings :

1. Statistic HU windowing: According to our observation, the HU values lying in the range of the window level plus or minus three standard deviations can portray most kidney areas and remove the outliers. However, in tumor detection, plus or minus two standard deviations obtain better results based on the characteristics of tumor or kidney images. In our study, the recommendation of various standard deviations is used to set different windowing ranges for the tumor and kidney detections.
2. Image screening: In the KiTS19 Challenge, the kidney-contained slices account for 36%, while non-kidney areas are 64% in the dataset. The tumor-contained slices are accounted for 12.57% of kidney-contained slices. The tumor segmentation aims to identify the areas. However, the tumors appear small-size in a tiny dataset for tumor detection. The results recommend putting all slices with tumor size greater than 0.01% to the training stages. A better dice score could be obtained than in other cases.
3. Label alternative: Following the best results of the above experiments, the tumors appear small-size, the semantic segmentation model concerning single or multiple targets possibly affects the identification accuracy. The HU windowing methods also influence the dice scores. In this case, the labeling way of the ground truth dataset obeys the following rules. The non-kidney areas are 0; the kidney areas denoted as 1 and 2 stand for kidney tumors in the KiTS19 Challenge. The windowing settings are both two standard deviations. The tumor labeling setting 0, 1, 2 with the slices with tumors and the tumor $>$ 0.01% are both better than the ones with {0, 0, 1}. The comparison diagram can be seen in Fig. 13.

4.3.6. Volume calculation

A typical application of kidney segmentation is the calculation of TKV. In clinical scenarios, doctors can estimate TKV by using the ellipsoid formula (6) [43]. However, because tumorous kidneys have irregular shapes, this equation is unsuitable for accurately calculating the TKV.

$$\text{TKV} = \text{Length} \times \text{Width} \times \text{Depth} \times \frac{\pi}{6} \tag{6}$$

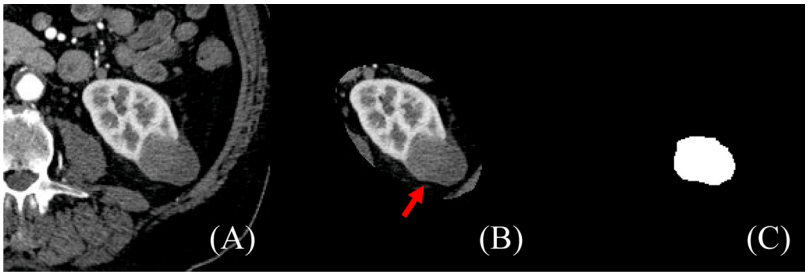

Fig. 9. (A) The original image (B) The image with the kidney mask (C) The estimated segmentation result of tumor region.

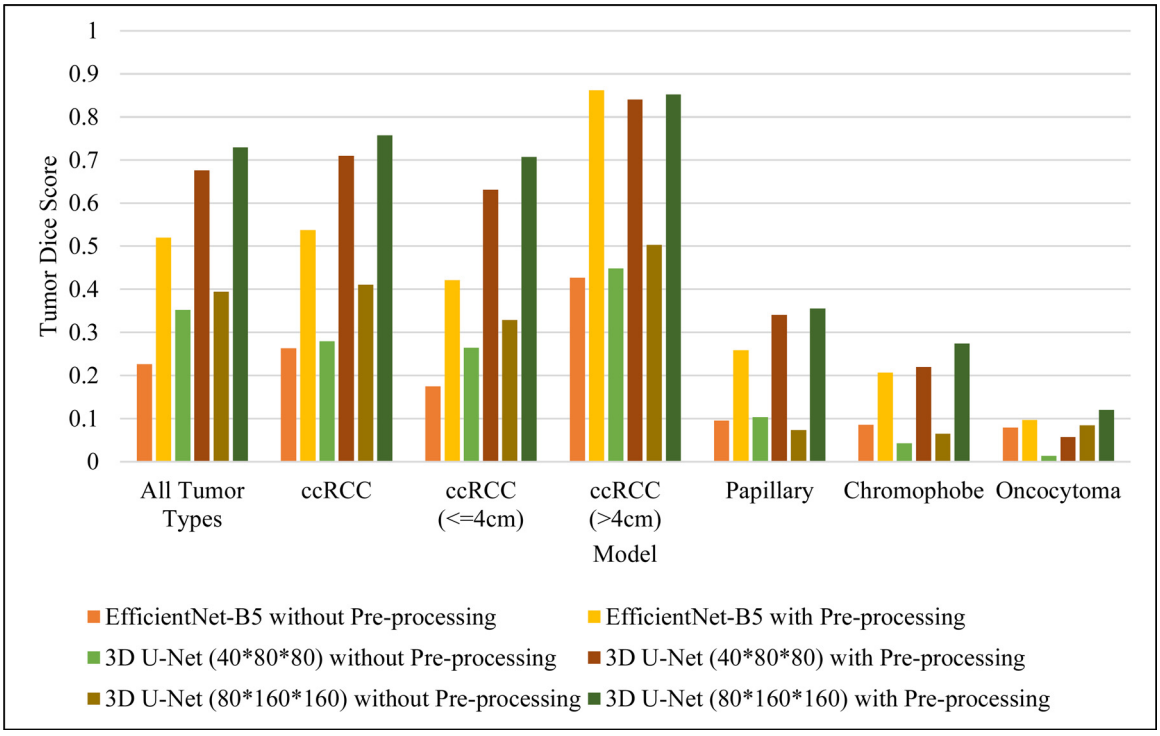

Fig. 10. The tumor segmentation performance is compared using various models for segmenting distinct tumor types in the KiTS datasets. The comparison also contains the KiTS datasets with and without pre-processing procedures.

Table 15
Comparison between ellipsoid formula calculator and our calculation methods. |AVG| indicates the average of the absolute values of the volume error for each case.

Case	Ground truth (ml)	Calculator (ml)	Volume error(%)	Ours (ml)	Volume error(%)
200	562.43	491	−12.7	563.64	0.2
201	608.21	563	−7.4	610.35	0.4
202	482.99	572	18.4	490.96	1.7
203	192.49	164	−14.8	195.06	1.3
204	529	433	−18.1	526.61	−0.5
205	331.09	339	2.4	329.46	−0.5
206	466.18	401	−14	446.79	−4.2
207	484.03	456	−5.8	487.96	0.8
208	398.45	402	0.9	379.64	−4.7
209	479.41	430	−10.3	479.52	0.02
|AVG|			10.48		**1.43**

Table 15 summarizes a comparison of TKV calculation methods. The proposed automatic TKV calculation method had an average volume error of only 1.43% from the ground truth, which was approximately seven times lower than the volume error obtained using conventional volume calculations. As shown in Table 16 and Fig. 14, our method achieved the state-of-the-art performance for TKV calculation.

5. Discussion

In our previous work, the transfer learning technique was applied to measure total kidney volume [3]. The Liver Tumor Segmentation (LiTS) Challenge dataset [34] and a private dataset (NTUH) were the source datasets. The two datasets are close to the KiTS datasets with similar pre-processing procedures. There-

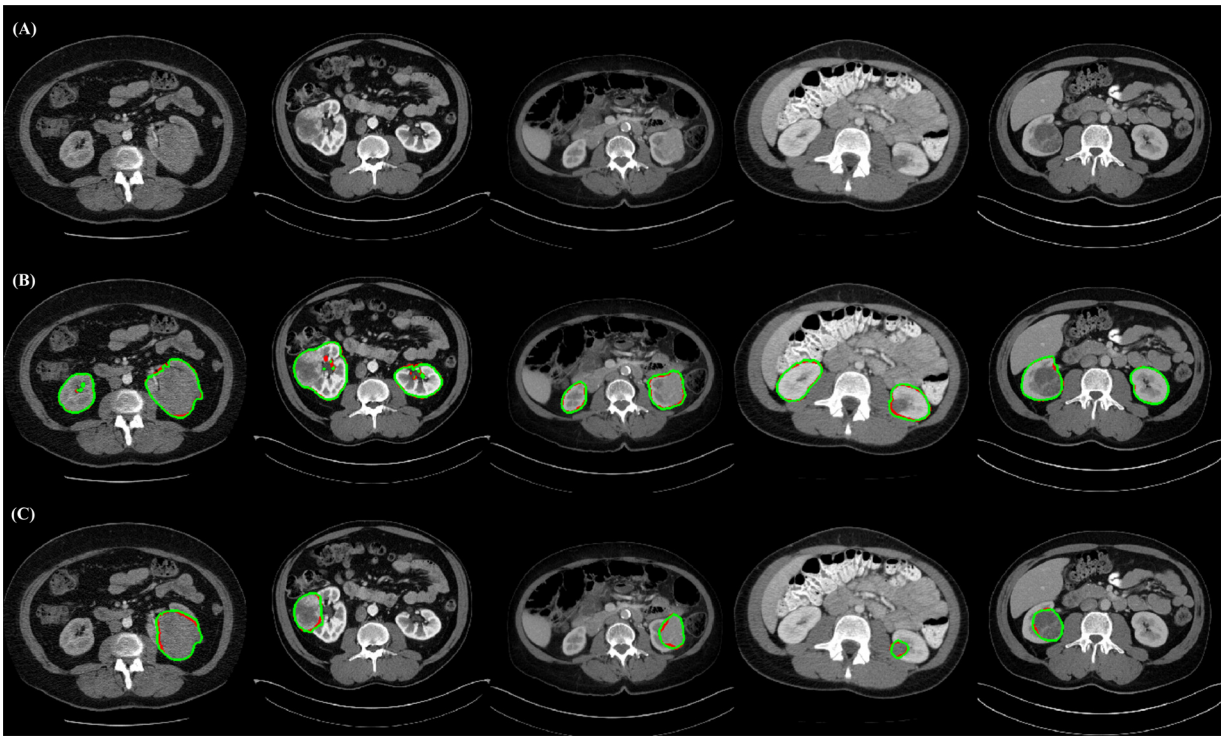

Fig. 11. The visual results for kidney and tumor segmentation by ResNet-41 on the pre-processing datasets: (A) Original image, (B) Kidney segmentation, (C) Tumor segmentation. The red line represents ground truth; the green line represents the segmented results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

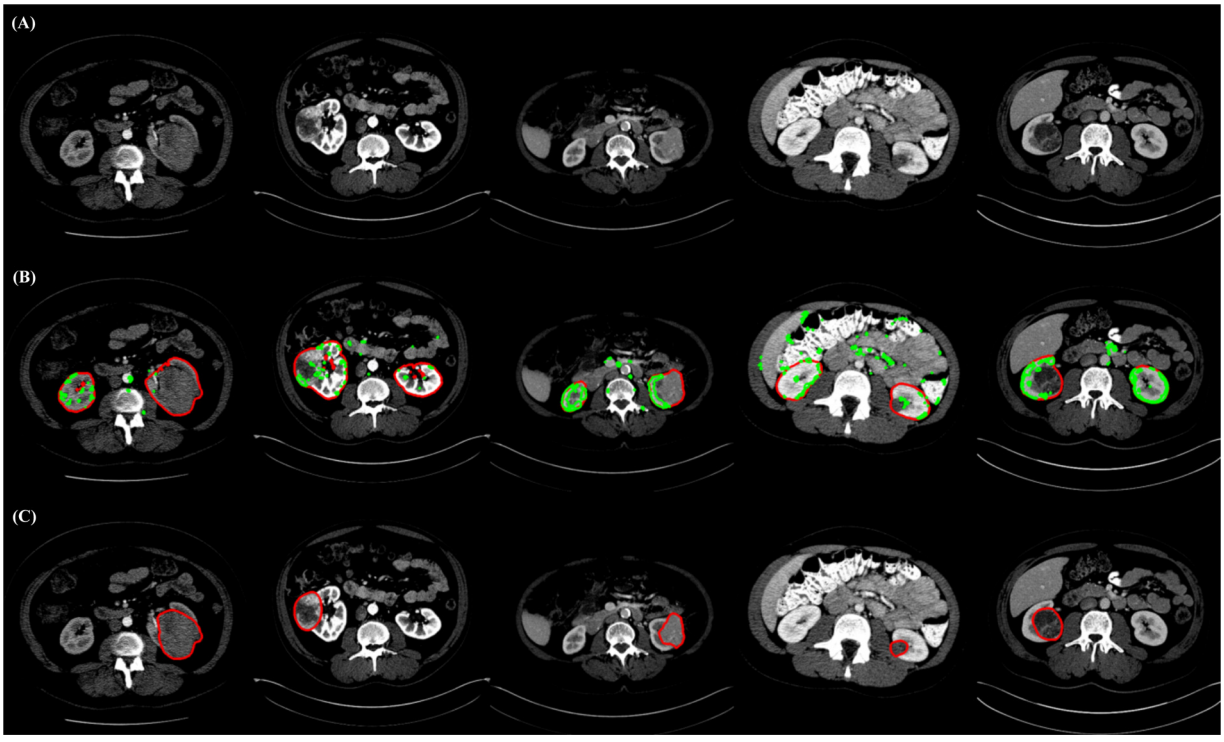

Fig. 12. The visual results for kidney and tumor segmentation by frozen encoder EfficientNet-B7 on the datasets without pre-processing : (A) Original image, (B) Kidney segmentation, (C) Tumor segmentation. The red line represents ground truth; the green line represents the segmented results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

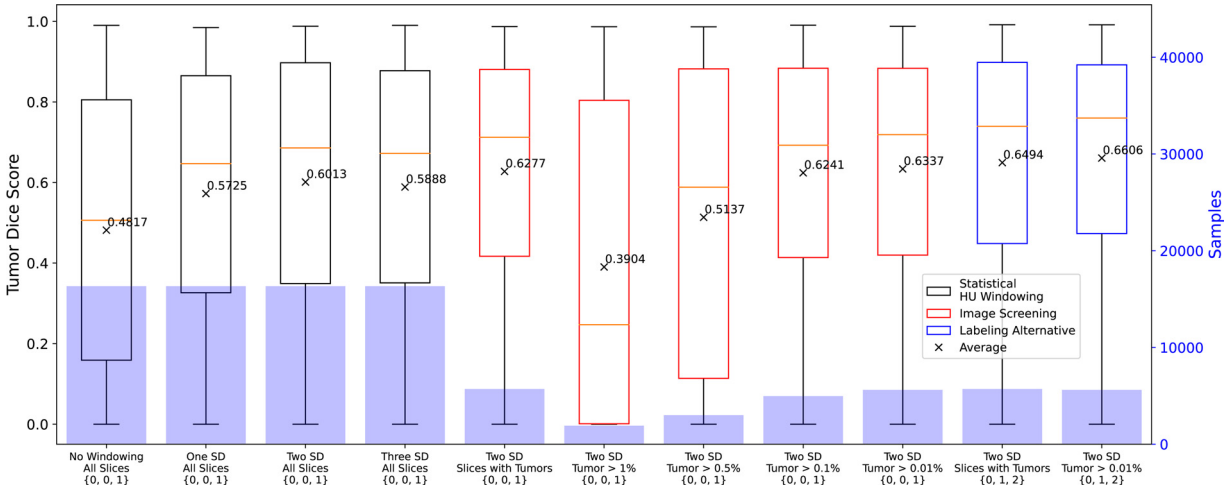

Fig. 13. The comparison of tumor segmentation results is obtained using with and without pre-processing methods with five-fold cross-validation. The most right case demonstrates the best performance.

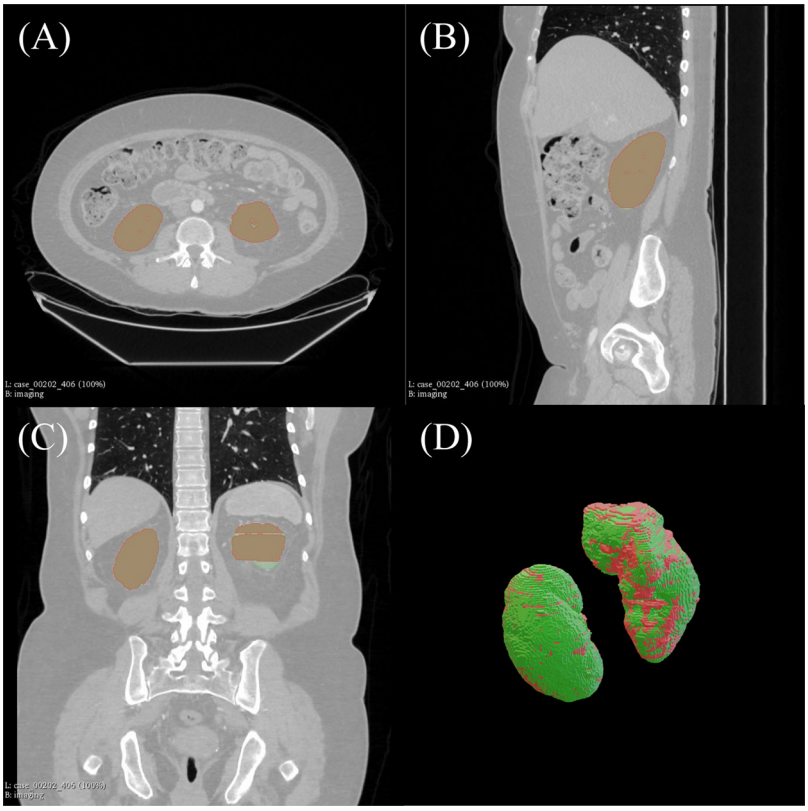

Fig. 14. Kidney segmentation results with the combined ResNet-41 and U-Net model; (A) Axial plane, (B) Sagittal plane, (C) Coronal plane, and (D) 3D demonstration. The red region denotes the prediction, and the green region denotes the ground truth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

fore, the pre-processing methods were systematically proposed to improve our study's current kidney and tumor segmentation methods, including windowing, unifying image direction, and particular case filtering. They can be a practical solution when annotated data are limited. The KiTS dataset was used as the target dataset to enhance the segmentation performance. The liver dataset was used as the source because liver segmentation, which focuses only on one region, is more straightforward than kidney segmentation is. The proposed model can accurately detect organ features. The results revealed that kidney segmentation accuracy was significantly improved by windowing methods, model selection, and transfer learning techniques to inspire us to make more information described in this paper.

Most current studies have focused on model adjustment, as discussed in Section 3.4. In the present study, we found that adjusting the dataset can also significantly improve segmentation performance. We performed windowing for each CT image and found that doing so within three standard deviations can enhance the performance compared with windowing within a range of one standard deviation. Furthermore, the distribution of the CT images may influence the performance. We found that all images, including those with and without kidney areas, should be used to train

Table 16
Comparison with existing methods of total kidney volume calculation.

	Methods	Targets	Volume error(%)
[44]	Manual ellipsoid fitting	Normal kidneys	14.2
[24]	Deep learning (nnUNet)	KiTS19	~6.74
[26]	Deep learning (VB-Net)	KiTS19	~5.58
[29]	Regression Forest	CKD	7.01
[45]	Deep learning (FCN)	KiTS19	4.8
This paper	Deep learning (Res-UNet)	KiTS19	**1.43**

the model. An optimal ratio of each type of image in the training dataset might be identifiable. Furthermore, using only images with kidney areas degraded the performance. With different image annotation methods, a label such as {background, kidney, kidney tumor} corresponding to {0, 1, 1} can reduce the noise and improve the segmentation performance.

In this study, we also investigated distinct architecture encoder models. The results show that EfficientNet-B7 with frozen parameters outperformed the other models. We froze the parameters of EfficientNet-B7 because of the computational limitations. However, the experiments with the EfficientNet-B4 model suggested that unfreezing the encoder's weights improves the performance. Therefore, we expect that EfficientNet-B7 with unfrozen weights can improve the performance. However, more computing resources (high-end graphic cards) are required.

Tumor segmentation is not as accurate as kidney segmentation, especially when the tumor is small. For tumors smaller than 4 cm, 3D U-Net resulted in a dice score of only 0.71, whereas for giant tumors, the dice score was 0.85. This difference may contribute to the class imbalance or inconsistent image resolution. Further, insufficient training data (KiTS19) were available for kidney tumor types other than ccRCC; therefore, kidney regions could not be well detected for these tumor types. Nevertheless, kidney volume can be calculated precisely.

6. Conclusion

This study systematically investigated medical image processing methods and discussed how they influenced segmentation performance. Specifically, we studied image windowing, image filtering, and image annotation definition. After these processes were performed step-by-step, the dice score improved from 0.9436 to 0.9524. For model adjustment, we applied the state-of-the-art EfficientNet model; as a result, the dice score further increased to 0.9648. This study used 3D U-Net to achieve a dice score of 0.7294 for all types of tumors. Under a primary computational environment, we proposed several medical image processing methods and used a state-of-the-art encoder-decoder models to predict kidneys and tumors. Eventually, we achieved a satisfactory performance with low computational burden. To accurately calculate TKV, this study also developed a deep learning–based kidney segmentation algorithm. The results achieved cost efficiency and the automatic volume calculator yielded a volume difference of only 1.43% compared with the ground truth.

Data availability

Datasets related to this article can be found at https://github.com/neheller/kits19, an open-source online data repository hosted by MICCAI 2019 [46].

Declaration of Competing Interest

Authors declare that they have no conflict of interest.

Acknowledgments

This study was supported by grants from the National Taiwan University Hospital Yunlin Branch (NTUHYL107.I0-07, NTUHYL109.I008 and NTUHYL 110.I004). The authors are grateful to National Taiwan University Hospital Yunlin Branch for assistance with the statistical analysis.

References

[1] R. Magistroni, C. Corsi, T. Mart, R. Torra, A review of the imaging techniques for measuring kidney and cyst volume in establishing autosomal dominant polycystic kidney disease progression, Am. J. Nephrol. 48 (1) (2018) 67–78 https://doi.org/10.1159/000491022.
[2] N. Tangri, I. Hougen, A. Alam, R. Perrone, P. MaFarlane, Y. Pei, Total kidney volume as a biomarker of disease progression in autosomal dominant polycystic kidney disease, Can. J. Kidney Health Dis. 4 (2017) 1–6, doi:10.1177/2054358117693355.
[3] C.-H. Hsiao, M.-C. Tsai, F.Y.-S. Lin, P.-C. Lin, F.-J. Yang, S.-Y. Yang, S.-Y. Wang, P.-R. Liu, Y. Huang, Automatic kidney volume estimation system using transfer learning techniques, in: L. Barolli, I. Woungang, T. Enokido (Eds.), Advanced Information Networking and Applications, Springer International Publishing, Cham, 2021, pp. 370–381.
[4] N. Zeng, H. Li, Y. Peng, A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease, Neural Comput. Appl. (2021) 1–12, doi:10.1007/s00521-021-06149-6.
[5] G. Litjens, T. Kooi, B.E. Bejnordi, A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. van Ginneken, C.I. Snchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, doi:10.1016/j.media.2017.07.005.
[6] L.B. da Cruz, J.D.L. Arajo, J.L. Ferreira, J.O.B. Diniz, A.C. Silva, J.D.S. de Almeida, A.C. de Paiva, M. Gattass, Kidney segmentation from computed tomography images using deep neural network, Comput. Biol. Med. 123 (2020) 103906–103918, doi:10.1016/j.compbiomed.2020.103906.
[7] T. Les, T. Markiewicz, M. Dziekiewicz, M. Lorent, Automatic recognition of the kidney in CT images, in: 19th International Conference Computational Problems of Electrical Engineering (CPEE), 2018, pp. 1–4, doi:10.1109/CPEE.2018.8506777.
[8] R. Kaur, M. Juneja, A survey of kidney segmentation techniques in CT images, Curr. Med. Imaging 14 (2) (2018) 238–250, doi:10.2174/1573405613666161221164146.
[9] K. Sharma, C. Rupprecht, A. Caroli, M.C. Aparicio, A. Remuzzi, M. Baust, N. Navab, Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease, Sci. Rep. 7 (1) (2017) 2049–2059, doi:10.1038/s41598-017-01779-0.
[10] K.B. de Raad, K.A. van Garderen, M. Smits, S.R. van der Voort, F. Incekara, E.H.G. Oei, J. Hirvasniemi, S. Klein, M. Starmans, The effect of preprocessing on convolutional neural networks for medical image segmentation, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 655–658, doi:10.1109/ISBI48211.2021.9433952.
[11] N. Zeng, Z. Wang, H. Zhang, K.-E. Kim, Y. Li, X. Liu, An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips, IEEE Trans. Nanotechnol. 18 (2019) 819–829, doi:10.1109/TNANO.2019.2932271.
[12] N. Zeng, H. Li, Z. Wang, W. Liu, S. Liu, F.E. Alsaadi, X. Liu, Deep-reinforcement-learning-based images segmentation for quantitative analysis of gold immunochromatographic strip, Neurocomputing 425 (2021) 173–180, doi:10.1016/j.neucom.2020.04.001.
[13] S. Perumal, T. Velmurugan, Preprocessing by contrast enhancement techniques for medical images, Int. J. Pure Appl. Math. 118 (18) (2018) 3681–3688. http://www.ijpam.eu
[14] D.R. Sarvamangala, R.V. Kulkarni, Convolutional neural networks in medical image understanding: a survey, Evol. Intell. (2021) 1–22, doi:10.1007/s12065-020-00540-3.
[15] KiTS19 Challenge Homepage, 2019, (https://kits19.grand-challenge.org/). Accessed: 2022-03-22.
[16] K.D. Toennies, Guide to Medical Image Analysis, Springer London, 2017.
[17] P.F. Christ, M.E.A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W.H. Sommer, S. Ahmadi, B.H. Menze, Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3d conditional random fields, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2016, pp. 415–423, doi:10.1007/978-3-319-46723-8_48.
[18] Y. Zhang, C. Zhong, Y. Zhang, Z. Shi, Z. He, Semantic Feature Attention Network for Liver Tumor Segmentation in Large-scale CT Database, 2019. https://arxiv.org/abs/1911.00282.
[19] X. Chen, R. Zhang, P. Yan, Feature fusion encoder decoder network for automatic liver lesion segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI), 2019, pp. 430–433, doi:10.1109/ISBI.2019.8759555.
[20] A. Ravishankar, S. Anusha, H.K. Akshatha, A. Raj, S. Jahnavi, J. Madhura, A Survey on noise reduction techniques in medical images, in: 2017 International

conference of Electronics, Communication and Aerospace Technology (ICECA), vol. 1, IEEE, 2017, pp. 385–389, doi:10.1109/ICECA.2017.8203711.
[21] M. Islam, K.N. Khan, M.S. Khan, Evaluation of preprocessing techniques for U-Net based automated liver segmentation, in: 2021 International Conference on Artificial Intelligence (ICAI), IEEE, 2021, pp. 187–192, doi:10.1109/ICAI52203.2021.9445204.
[22] H. Badakhshannoory, P. Saeedi, A model-based validation scheme for organ segmentation in CT scan volumes, IEEE Trans. Biomed. Eng. 58 (9) (2011) 2681–2693, doi:10.1109/TBME.2011.2161987.
[23] F. Isensee, K.H. Maier-Hein, An Attempt at Beating the 3D U-Net, 2019, arXiv:1908.02182.
[24] X. Hou, C. Xie, F. Li, J. Wang, C. Lv, G. Xie, Y. Nan, A triple-stage self-guided network for kidney tumor segmentation, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 341–344, doi:10.1109/ISBI45749.2020.9098609.
[25] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241, doi:10.1007/978-3-319-24574-4_28.
[26] G. Mu, Z. Lin, M. Han, G. Yao, Y. Gao, Segmentation of kidney tumor by multi-resolution VB-nets, 2019, http://results.kits-challenge.org/miccai2019/manuscripts/gr_6e.pdf.
[27] F. Milletari, N. Navab, S.A. Ahmadi, V-Net: fully convolutional neural networks for volumetric medical image segmentationn, in: 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 565–571, doi:10.1109/3DV.2016.79.
[28] X.F. Xi, L. Wang, V.S. Sheng, Z. Cui, B. Fu, F. Hu, Cascade U-ResNets for simultaneous liver and lesion segmentation, IEEE Access 8 (2020) 68944–68952, doi:10.1109/ACCESS.2020.2985671.
[29] M.A. Hussain, G. Hamarneh, R. Garbi, Cascaded regression neural nets for kidney localization and segmentation-free volume estimation, IEEE Trans. Med. Imaging 40 (6) (2021) 1555–1567, doi:10.1109/TMI.2021.3060465.
[30] B. Baheti, S. Innani, S. Gajre, S. Talbar, Eff-UNet: a novel architecture for semantic segmentation in unstructured environment, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 1473–1481, doi:10.1109/CVPRW50498.2020.00187.
[31] L.T.T. Hong, N.C. Thanh, T.Q. Long, Polyp segmentation in colonoscopy images using ensembles of U-Nets with EfficientNet and asymmetric similarity loss function, in: 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), 2020, pp. 1–6, doi:10.1109/RIVF48685.2020.9140793.
[32] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med. Imaging 39 (6) (2020) 1856–1867, doi:10.1109/TMI.2019.2959609.
[33] X. Yan, K. Yuan, W. Zhao, S. Wang, Z. Li, S. Cui, An efficient hybrid model for kidney tumor segmentation in CT images, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 333–336, doi:10.1109/ISBI45749.2020.9098325.
[34] Liver Tumor Segmentation Challenge, 2017, (https://competitions.codalab.org/competitions/17094). Accessed: 2021-01-14.
[35] S.P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, B. Gulys, 3D deep learning on medical images: a review, Sensors 20 (18) (2020) 1–24, doi:10.3390/s20185097.
[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi:10.1109/CVPR.2016.90.
[37] M. Tan, Q. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 97, PMLR, 2019, pp. 6105–6114. https://proceedings.mlr.press/v97/tan19a.html
[38] T.-Y. Lin, P. Dollr, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 936–944, doi:10.1109/CVPR.2017.106.
[39] Q. Xie, M.-T. Luong, E. Hovy, Q.V. Le, Self-training with noisy student improves ImageNet classification, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10684–10695, doi:10.1109/CVPR42600.2020.01070.
[40] R.V. Hogg, E.A. Tanis, D.L. Zimmerman, Probability and Statistical Inference, vol. 993, Macmillan New York, 1977.
[41] V. Ficarra, G. Novara, A. Galfano, G. Novella, D. Schiavone, W. Artibani, Application of TNM, 2002 version, in localized renal cell carcinoma: is it able to predict different cancer-Specific survival probability? Urology 63 (6) (2004) 1050–1054, doi:10.1016/j.urology.2004.01.024.
[42] S.M. Nazim, M.H. Ather, K. Hafeez, B. Salam, Accuracy of multidetector CT scans in staging of renal carcinoma, Int. J. Surg. 9 (1) (2011) 86–90, doi:10.1016/j.ijsu.2010.07.304.
[43] M.V. Irazabal, L.J. Rangel, E.J. Bergstralh, S.L. Osborn, A.J. Harmon, J.L. Sundsbak, K.T. Bae, A.B. Chapman, J.J. Grantham, M. Mrug, M.C. Hogan, Z.M. El-Zoghby, P.C. Harris, B.J. Erickson, B.F. King, V.E. Torres, the CRISP Investigators, Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials, J. Am. Soc. Nephrol. 26 (1) (2015) 160–172, doi:10.1681/ASN.2013101138.
[44] N. Zakhari, B. Blew, W. Shabana, Simplified method to measure renal volume: the best correction factor for the ellipsoid formula volume calculation in pretransplant computed tomographic live donor, Urology 83 (6) (2015) e1444.e15–e1444.e19, doi:10.1016/j.urology.2014.03.005.
[45] M.A. Hussain, G. Hamarneh, T.W. O'Connell, M.F. Mohammed, R. Abugharbieh, Segmentation-free estimation of kidney volumes in CT with dual regression forests, in: L. Wang, E. Adeli, Q. Wang, Y. Shi, H.-I. Suk (Eds.), Machine Learning in Medical Imaging, 2016, pp. 156–163, doi:10.1007/978-3-319-47157-0_19.
[46] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, C. Weight, The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes, 2020. arXiv:1904.00445