



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이학석사 학위논문

**Measurement of Glomerular Filtration
Rate using Quantitative SPECT/CT and
Deep-learning-based Kidney
Segmentation**

정량적 SPECT/CT와 딥러닝 기반의 신장 분할을
이용한 사구체여과율 측정 방법

2019년 2월

서울대학교 대학원
의과학과

박 준 영

Master's Thesis

**Measurement of Glomerular Filtration
Rate using Quantitative SPECT/CT
and Deep-learning-based Kidney
Segmentation**

정량적 SPECT/CT와 딥러닝 기반의 신장 분할을
이용한 사구체여과율 측정 방법

February 2019

**Department of Biomedical Sciences,
The Graduate School
Seoul National University**

Jun Young Park

정량적 SPECT/CT와 딥러닝 기반의 신장 분할을 이용한 사구체여 과율 측정 방법

지도 교수 이재성

이 논문을 이학석사 학위논문으로 제출함

2019 년 2 월

서울대학교 대학원
의과학과

박 준 영

박준영의 이학석사 학위논문을 인준함

2019 년 2 월

위 원 장 _____ (인)
부위원장 _____ (인)
위 원 _____ (인)

Measurement of Glomerular Filtration Rate using Quantitative SPECT/CT and Deep-learning-based Kidney Segmentation

Examiner Jae Sung Lee

Submitting a master's thesis of Public Administration

February 2018

**Department of Biomedical Sciences,
The Graduate School
Seoul National University**

Jun Young Park

Confirming the master's thesis written by

Jun Young Park

February 2018

Chair _____(Seal)

Vice Chair _____(Seal)

Examiner _____(Seal)

Abstract

Measurement of Glomerular Filtration Rate using Quantitative SPECT/CT and Deep-learning-based Kidney Segmentation

Jun Young Park
Department of Biomedical Sciences,
The Graduate School
Seoul National University

Quantitative SPECT/CT is potentially useful for more accurate and reliable measurement of glomerular filtration rate (GFR) than conventional planar scintigraphy. However, manual drawing of a volume of interest (VOI) on renal parenchyma in CT images is a labor-intensive and time-consuming task. The aim of this study is to develop a fully automated GFR quantification method based on a deep learning approach to the 3D segmentation of kidney parenchyma in CT.

Totally, 393 patients underwent quantitative ^{99m}Tc -DTPA SPECT/CT scans. One-minute SPECT data were acquired 2 min after the intravenous tracer injection. Nuclear medicine physicians drew VOIs on renal parenchyma using the vendor's software. A modified 3D U-Net that learns an end-to-end mapping between CT and renal parenchyma segmented volumes was used. For

quantitative performance evaluation, the Dice similarity coefficient between manual drawing and deep learning output was calculated. I calculated the percentage injected dose (%ID) by applying the deep learning output (3D VOI) to the quantitative SPECT images, and assessed the correlation between the measurement of GFR using both segmentation methods. Clinical validation of automatic segmentation was also done in urolithiasis patients (n = 63) and kidney donors (n = 25). Totally, 176 kidneys were classified into three groups depending on the presence, location (ureter or renal), and size (cutoff size: 10 mm) of the radio-opaque stone: normal, asymptomatic, or symptomatic kidneys. The performance of the network was evaluated considering total GFR and individual GFR values in different groups of patients and kidneys.

I automatically segmented the kidneys in CT images using the proposed method with remarkably high Dice similarity coefficient relative to the manual segmentation (mean = 0.89). The GFR values derived using manual and automatic segmentation methods were strongly correlated ($R^2 = 0.96$). The absolute difference between the individual GFR values using manual and automatic methods was only 2.90%. Moreover, the two segmentation methods had comparable performance in the urolithiasis patients and kidney donors. Furthermore, both segmentation modalities showed significantly decreased individual GFR in symptomatic kidneys compared with the normal or asymptomatic kidney groups.

The proposed approach enables fast and accurate GFR measurement.

Keyword : deep learning; kidney parenchyma segmentation; glomerular filtration rate; single-photon emission computed tomography; computed tomography

Student Number : 2017-24168

Table of Contents

Chapter 1. Introduction	7
1.1. Study Background	7
1.2. Purpose of Research	8
Chapter 2. Materials and Methods	10
2.1. Dataset	10
2.2. Neural Network Architecture.....	13
2.3. Network Training	14
2.4. GFR Estimation	15
2.5. Further Validation on Urolithiasis Patients.....	15
2.6. Data Analysis	16
Chapter 3. Results	18
3.1. Segmentation	18
3.2. GFR Estimation	22
3.3. Validation on Urolithiasis Patients	23
Chapter 4. Discussion	28
Chapter 5. Conclusion	31
Bibliography	32
Abstract in Korean	35

List of Figures

Figure 1. Schematic diagrams of the deep-learning-based renal parenchyma segmentation for the measurement of glomerular filtration rate (GFR) using quantitative single-photon emission computed tomography (SPECT)/computed tomography (CT). 1 2

Figure 2. Deep neural network architecture. The network learns an end-to-end mapping between computed tomography (CT) and renal parenchyma segmented volumes. The network consists of the contraction and expansion paths. 1 3

Figure 3. SPECT/CT images and renal parenchyma VOIs with incorrect ROI interpolation result (next slice as the second and fourth column) provided from the vendor’s software. (A) Manually segmented VOI. (B) Deep-learning-generated automatic VOI. 1 9

Figure 4. SPECT/CT images (multiple renal stones for (A), (B) and partial nephrectomy for (C), (D)) and renal parenchyma VOIs for a representative test dataset (A), (C) Manually segmented VOI. (B), (D) Deep-learning-generated automatic VOI. 2 1

Figure 5. Scattered (A-C) and Bland–Altman (D-F) plots between measurement of left glomerular filtration rate (GFR) using manual and deep-learning-generated volumes of interest (VOIs). (A), (C) Total kidney. (B), (E) Left kidney. (C), (F) Right kidney. 2 2

Figure 6. Absolute percentage difference between measurement of GFR using manual and deep-learning-generated VOIs: results of five-fold cross-validation. 2 3

Figure 7. Total and Individual GFR comparison. (A) Total GFR comparison of manual and CNN-based automatic segmentation methods in kidney donor (normal) and urinary stone patients (stone). (B) Individual GFR comparison of normal, asymptomatic (Asym), or symptomatic (Sym) kidneys in manual

and automatic segmentation methods. *P < 0.001..... 2 4

Figure 8. SPECT/CT images of a patient (A) before (red arrow indicates a ureter stone, yellow arrow indicates a renal stone) and (B) 4 months after removal of left ureter and renal stones. A projection image is presented in the first column, axial images of CT (top) and SPECT/CT fusion (bottom) are in the second, and segmentation results (automatic segmentation in the top and manual segmentation in the bottom) are shown in the third column.
..... 2 6

Figure 9. SPECT/CT images of a patient (A) before (red arrow indicates a ureter stone, yellow arrow indicates a renal stone) and (B) 4 months after removal of left ureter and renal stones. A projection image is presented in the first column, axial images of CT (top) and SPECT/CT fusion (bottom) are in the second, and segmentation results (automatic segmentation in the top and manual segmentation in the bottom) are shown in the third column.
..... 2 8

Figure 10. SPECT/CT images (renal mass and renal pelvis included) and renal parenchyma VOIs for a representative test dataset. (A) Manually segmented VOI. (B) Deep-learning-generated automatic VOI. 2 9

Figure 11. SPECT/CT images (single kidney) and renal parenchyma volumes of interest (VOIs) for a representative test dataset. (A) Manually segmented VOI. (B) Deep-learning-generated automatic VOI. 2 9

List of Tables

Table 1. Demographics of patients for network training and validation (n = 393)	1 0
Table 2. The results of cross-validations (total kidney).....	1 8
Table 3. Ablation study between conventional 3D U-Net and proposed network	1 9
Table 4. Total GFR (ml/min/1.73 m ²) by manual and CNN-based segmentations in normal and urolithiasis patients (mean \pm SD)....	2 4
Table 5. Individual GFR (ml/min) by manual and CNN-based segmentations in normal, symptomatic, and asymptomatic urinary stone kidneys (mean \pm SD).....	2 5
Table 6. Change of individual kidney GFR before and after stone removal procedure in a urolithiasis patient.....	2 6

Chapter 1. Introduction

1.1. Study Background

Glomerular filtration rate (GFR) is defined as a flow rate of blood plasma that is filtered through glomerulus. It is considered as an indicator of renal function and is routinely used to stratify the severity of acute kidney injury or chronic kidney disease. The GFR is estimated using a substance that is completely filtered through glomerulus, not reabsorbed and not excreted in renal tubules (1). If that is the case, urinary clearance of the substance is equal to the plasma clearance. Formulas such as Modification of Diet in Renal Disease (MDRD) or Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equations are frequently used in the clinic to derive GFR from the serum creatinine level, although creatinine is not an ideal substance for measuring renal clearance (2).

In nuclear medicine, ^{51}Cr -ethylenediaminetetraacetic acid (EDTA) and $^{99\text{m}}\text{Tc}$ -diethylenetriaminepentaacetic acid (DTPA) are the two most commonly utilized radiopharmaceuticals to evaluate renal function. Blood or urine sampling after injection of the radiotracers is a method to directly measure renal clearance (1,3). However, these sampling procedures are laborious and time-consuming in comparison with other approaches. Therefore, $^{99\text{m}}\text{Tc}$ -DTPA planar scintigraphy that measures the radiation counts in each kidney is more commonly used because of its ease of use. The most popular method to calculate the GFR from the radiation counts is Gate's method that utilizes the counts measured for 1 min from 2 min to 3 min after the injection of $^{99\text{m}}\text{Tc}$ -DTPA (4,5).

Kidney single-photon emission computed tomography (SPECT)/computed tomography (CT) with $^{99\text{m}}\text{Tc}$ -DTPA is a promising method for the measurement of

GFR because it is more quantitative and reliable in measuring the renal clearance than planar scintigraphy (6). The SPECT/CT is acquired during the same period of time as in the renal planar scintigraphy, and the volume of interest (VOI) is drawn on each kidney for the quantification of absolute radioactivity. The measurement of GFR using ^{99m}Tc -DTPA SPECT/CT was reproducible and accurate in healthy volunteers and renal tumor patients post partial nephrectomy and useful for disease severity evaluation in urinary stone patients (6,7). However, the necessity of manual drawing of the VOI on the whole renal parenchyma in CT images is an obstacle that prevents the wide use of this new approach. The labor-intensive and time-consuming manual drawing usually takes about 15 min per scan by nuclear medicine physicians (6,7).

1.2. Purpose of Research

The aim of this study is to develop an automated GFR quantification method based on deep learning approach to the three-dimensional (3D) segmentation of kidney parenchyma in CT acquired in quantitative kidney SPECT/CT studies. In recent years, deep convolutional neural networks (CNNs) have shown superior performance in many computer vision and biomedical applications, such as image de-noising, image classification and object detection, and organ segmentation (8-18). In addition, there is a related deep-learning-based kidney segmentation study for total kidney volume (functioning parenchyma and non-functioning cysts) quantification in autosomal dominant polycystic kidney disease (ADPKD) (19). In addition, some studies have shown that deep-learning based 3D segmentation is effective for medical image dataset (20-23). However, the GFR quantification

needs the segmentation of the only functioning kidney parenchyma, which is more sophisticated than total kidney segmentation in ADPKD. In this study, I have trained a deep CNN to learn end-to-end mapping between the 3D CT volume and manually segmented VOI by experts using a dataset including 315 patients. The performance of the CNN was validated using another dataset including 78 patients, and five-fold cross-validation was performed. Finally, the measurement of GFR using the manually drawing and deep-learning-generated VOIs were compared with each other in 63 urolithiasis patients and 25 negative controls (kidney donor) to show the clinical validity of the proposed method.

Chapter 2. Materials and Methods

2.1. Dataset

Quantitative ^{99m}Tc -DTPA kidney SPECT/CT data of 393 patients (257 men and 136 women, age = 53.55 ± 12.64 years) were retrospectively analyzed for network training and validation (**Table 1**).

Table 1. Demographics of patients for network training and validation (n = 393)

Patient Characteristics	Data	
Gender (male:female)	257:136	
Age (years)	53.55 ± 12.64	
Height (cm)	166.04 ± 9.23	
Weight (kg)	70.17 ± 12.86	
Body surface area* (m ²)	1.78 ± 0.19	
Serum Creatinine (mg/dl)	0.89 ± 0.26	
eGFR† (ml/min/1.73 m ²)	90.63 ± 18.18	
Disease	Normal (kidney donor)	18
	Renal tumor	83
	Urinary stone	112
	Post partial nephrectomy‡	174
	Post total nephrectomy§	5
	Other	1
Contrast-enhanced CT (yes:no)	89:304	

* The equation for body surface area is the Dubois formula:

$$\text{Body surface area (m}^2\text{)} = 0.007184 \times (\text{weight in kg})^{0.425} \times (\text{height in cm})^{0.725}$$

† Estimated glomerular filtration rate (eGFR) is derived from the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation.

‡ Cause of partial nephrectomy: renal tumor (173) or simple cyst (1).

§ Cause of total nephrectomy: renal tumor (5)

Data=mean \pm standard deviation. CT, computed tomography.

The retrospective use of the scan data and waiver of consent were approved by the Institutional Review Board of my institute. The SPECT/CT data were acquired using a GE Discovery NM/CT 670 scanner equipped with a low-energy

high-resolution collimator (6). One-minute SPECT data were acquired in a continuous mode 2 min after the intravenous injection of 370 MBq ^{99m}Tc -DTPA. The peak energy was set at 140 KeV with 20% window (126–154 KeV), and the scatter energy was set at 120 KeV with 10% window (115–125 KeV). The SPECT images were reconstructed using an iterative ordered subset expectation maximization (OSEM) algorithm (2 iterations and 10 subsets) and corrected for attenuation, scatter, and collimator–detector response. A post-reconstruction low pass filter (Butterworth with frequency of 0.48 and order of 10) was applied and image matrices were $128 \times 128 \times 128$ (voxel size: 3.452 mm^3). The applied zoom factor during SPECT acquisition was 1.28. The CT acquisition conditions were as follows: tube voltage of 120 KVp, tube current of 60–210 mA with autoMA function at a noise level of 20, detector collimation of 20 mm ($=16 \times 1.25 \text{ mm}$), helical thickness of 2.5 mm, table speed of 37 mm/s, tube rotation time of 0.5 s, and pitch of 0.938:1. The CT images were reconstructed in a $512 \times 512 \times 161$ matrix with voxel sizes of $0.977 \times 0.977 \times 2.5 \text{ mm}^3$. The system sensitivity of SPECT for ^{99m}Tc was 152.5 cpm/ μCi , which had been determined by triple independent sessions of phantom studies (24).

A nuclear medicine physician manually drew 2D regions of interest (ROIs) on individual renal parenchyma. To cover the whole kidney volume, approximately 80–100 slices were required in normal kidney. To save time and effort, ROIs were drawn in every 2–3 coronal CT slices up to 30 slices using the vendor’s Q Metrix software, with the exclusion of unwanted structures like cysts, urinary stones, and tumors. After automatic ROI interpolation between the slices, which is provided by Q Metrix, a VOI was then generated by integrating these manually drawn and interpolated single-slice ROIs. While the automatic ROI interpolation is useful for

reducing the time and labor required for the VOI drawing, the interpolated slices may still include unwanted structures, requiring further manual interventions. It is of note that the quantitative kidney SPECT/CT was performed without iodine-contrast enhancement; however, iodine-contrast remained in the renal pelvis of the SPECT/CT in 22.6% (=89/393) because contrast-enhanced CT was performed 1–2 hours prior to the SPECT/CT for the post-operative evaluation of renal tumor (Table 1).

For CNN training and data analysis, CT and SPECT images and manual VOIs were resampled to have the same matrix and voxel sizes ($256 \times 256 \times 232$ and 1.726 mm^3). To reduce the memory consumption in CNN training, the images were then cropped into $192 \times 128 \times 96$ matrices, which are large enough to include both kidneys. In addition, I applied 3D volume smoothing and morphological operations to the manual VOIs to reduce the discontinuity in 3D space caused by the 2D ROI drawing. These preprocessed images and VOIs were finally used for CNN training and testing and GFR estimation as shown in **Figure 1**.

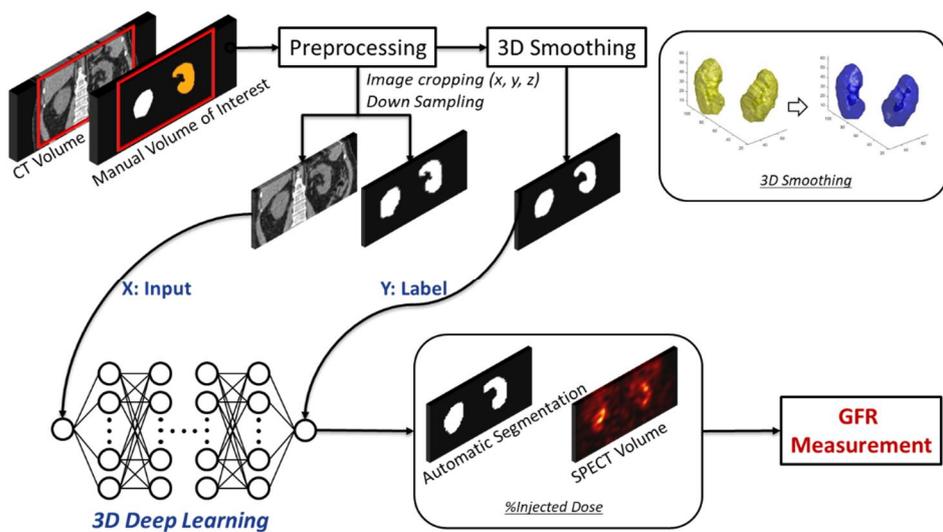


Figure 1. Schematic diagrams of the deep-learning-based renal parenchyma segmentation for the measurement of glomerular filtration rate (GFR) using

quantitative single-photon emission computed tomography (SPECT)/computed tomography (CT).

2.2. Neural Network Architecture

The CNN that I used is a modified 3D U-net that consists of the contraction and expansion paths (25). The 3D U-net learns an end-to-end mapping between CT and manually drawn renal parenchyma VOIs as shown in **Figure 2**.

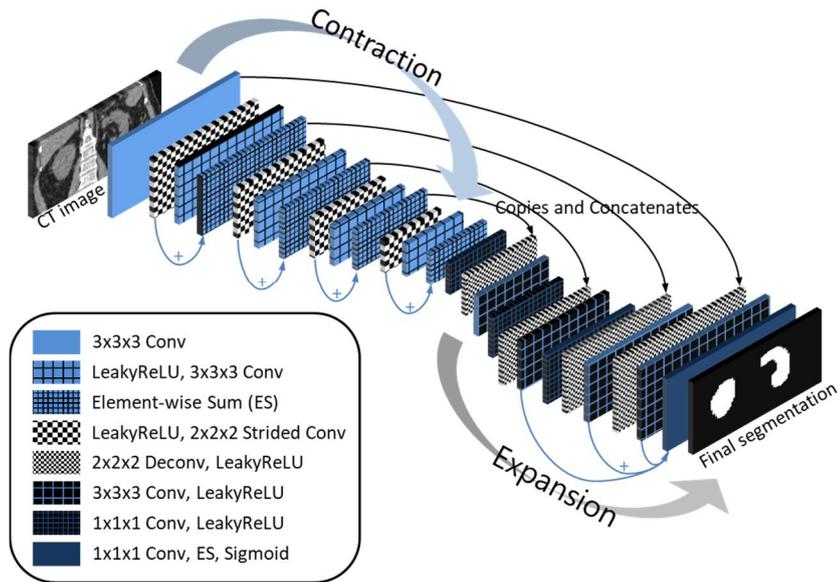


Figure 2. Deep neural network architecture. The network learns an end-to-end mapping between computed tomography (CT) and renal parenchyma segmented volumes. The network consists of the contraction and expansion paths.

Each path is exploited by the five sequential layers. The contraction path, which captures the context, consist of a leaky rectified linear unit (leaky ReLU) as an activation function (a pre-activation residual block (26)), each followed by $3 \times 3 \times 3$ convolution and $2 \times 2 \times 2$ strided convolution for down-sampling. In addition, the element-wise sum array is used between the output of $2 \times 2 \times 2$ strided convolution and $3 \times 3 \times 3$ convolution to forward feature maps from one stage of the network to other (20). The expansion path, which enables precise localization,

consists of the leaky ReLU, which allows a small gradient when the unit is not active, $3 \times 3 \times 3$ convolution, $1 \times 1 \times 1$ convolution, and $2 \times 2 \times 2$ de-convolution for up-sampling. The element-wise sum array layer is also used right before the Sigmoid activation function to sum $3 \times 3 \times 3$ convolution results of the previous three layers (20). We used a 3D spatial drop-out technique (drop-out rate of 0.3), as these have shown better performance when adjacent voxels within feature maps are strongly correlated compared with batch normalization.

I also employed symmetric skip connections (copy and concatenation), as shown in **Figure 2**, to insert the local details captured in the feature maps of the contraction path into the feature maps of the expansion path (27).

I implemented the networks using the TensorFlow (28) and Keras framework (<https://github.com/fchollet/keras>).

2.3. Network Training

The network was trained using randomly selected dataset of 315 of 393 patients and validated using data from the remaining 78 patients. As mentioned previously, the resampled and cropped CT images and manual VOIs were used for network training. All the input and output datasets were in 3D volume format.

The Dice similarity coefficient, which is an overlap metric frequently used for assessing the quality of segmentation maps, is used as the loss function (11,21,29). Each layer was updated using error back-propagation with adaptive moment estimation optimizer, which is a stochastic optimization technique (30). The exponential decay rates for the moment estimates β_1 and β_2 are 0.9 and 0.999 respectively, with epsilon of zero. The learning rate for determining to what extent

the newly acquired information overrides the old information was initially 0.0005 and reduced by half after 10 epochs if the loss function is not improved. The number of epochs was 80 and each epoch includes 272 iterations. The training time was approximately 60 min/epoch when using i7-7700K CPU (3.40 GHz) and one GTX 1080 TI GPU.

2.4. GFR Estimation

I calculated %ID by applying the manual and automatic VOIs to the quantitative SPECT images. Then, individual GFRs were calculated using the following equation (6):

$$\text{GFR (ml/min)} = (\%ID \times 9.1462) + 23.0653$$

The sum of bilateral kidney GFRs was normalized to body surface area (BSA) using the following equation to calculate the total GFR of the bilateral kidneys. The Dubois equation for the BSA in m² was $0.007184 \times (\text{weight in kg})^{0.425} \times (\text{height in cm})^{0.725}$.

$$\text{Total GFR (ml/min/1.73 m}^2\text{)} = \text{GFR (ml/min)} \times (1.73/\text{BSA m}^2\text{)}$$

2.5. Further Validation on Urolithiasis Patients

To evaluate the performance of the network in the clinical setting, I adopted patients with urinary stones and kidney donors as negative controls. Consecutive ^{99m}Tc-DTPA kidney SPECT/CT studies of urolithiasis patients or kidney donors performed from March 2015 to January 2016 were analyzed retrospectively. Among 69 urinary stone patients scanned during that period, 4 with underlying chronic kidney disease and 2 without available raw data were excluded. Among 26 kidney donors, one subject without raw data remaining was also excluded. Finally,

126 kidneys from 63 urinary stone patients and 50 kidneys from 25 kidney donors were investigated. Gender proportion was not significantly different between the normal (male:female = 15:10) and stone group (male:female = 30:33) (Chi-square test, $p = 0.086$). However, age was significantly higher in urinary stone subjects (56.87 ± 12.60 years old) than in normal subjects (45.64 ± 13.91 years old) (independent samples t -test, $p = 0.0004$).

Each kidney was classified into three groups: 50 normal kidneys (from kidney donor patients), 48 symptomatic kidneys with either ureter stone of any size or large renal stone (longest diameter >10 mm), and 78 asymptomatic kidneys with either small renal stone (longest diameter ≤ 10 mm) or contralateral kidney of unilateral urolithiasis patients.

Individual and total GFR values obtained from manual segmentation or automatic segmentation were compared in each group. For manual segmentation, the average of two independent measurements of GFR by four medical experts was used to represent the manual GFR. The experts were blind to each other regarding the manual segmentation results. For automatic segmentation, only a single measurement of GFR by the deep learning algorithm was employed.

2.6. Data Analysis

For the quantitative evaluation of network performance, the Dice similarity coefficient between manual drawing and deep learning output was calculated. In addition, I assessed the correlation and mean absolute percentage error between the measurements of GFR using these different segmentation methods. To confirm the consistency of performance, I also performed five-fold cross-validation.

Statistical analyses in urolithiasis patients were performed with dedicated software (Medcalc, version 14.8.1, bvba/GraphPad Prism, version 5.01). First, normality of the data was evaluated using the D'Agostino–Pearson test and parametric or non-parametric tests were implemented according to the result. For the parametric test, independent samples *t*-test, paired samples *t*-test, or one-way analysis of variance (ANOVA) was performed. For the non-parametric test, Mann–Whitney test, Wilcoxon test, or Kruskal–Wallis test was done. Chi-square tests were performed for analyses of categorical data. A multiple comparison correction for *t*-test was implemented with Bonferroni correction. Results with *P*-values less than 0.05 were considered significant.

Chapter 3. Results

3.1. Segmentation

I could automatically segment the kidneys in CT images using the proposed method with high Dice similarity coefficient relative to the manual segmentation (mean \pm SD = 0.89 ± 0.03 in main experiment) (**Table 2**).

Table 2. The results of cross-validations (total kidney)

Method	Unit	Dataset				
		Main Experiment	Cross-validation 1	Cross-validation 2	Cross-validation 3	Cross-validation 4
DSC	(mean \pm SD)	0.89 ± 0.03	0.88 ± 0.04	0.88 ± 0.04	0.89 ± 0.03	0.89 ± 0.03
	[range]	0.80–0.93	0.65–0.93	0.74–0.94	0.77–0.93	0.79–0.94
Mean-M	ml/min (mean \pm SD)	49.87 ± 0.08	49.28 ± 0.21	47.83 ± 9.58	50.27 ± 0.72	49.68 ± 0.06
Mean-A	ml/min (mean \pm SD)	49.41 ± 9.81	48.89 ± 9.88	47.70 ± 9.07	49.95 ± 0.26	49.06 ± 9.68
Co- rrelation	R^2	0.96	0.96	0.95	0.96	0.96
MAPE	% (mean \pm SD)	2.90 ± 2.80	2.88 ± 2.75	2.99 ± 3.25	3.00 ± 2.93	2.67 ± 2.70

DSC, Dice similarity coefficient; M, manual segmentation; A, automatic segmentation; MAPE, mean absolute percentage error.

In addition, the proposed deep learning approach provided 3D kidney parenchyma VOIs with no discontinuity between slices because the CNN was trained to produce smooth 3D VOIs. The time requirement of auto-segmentation was only a few seconds per patient, whereas the manual segmentation takes about 15 min per scan.

We also performed an ablation study to optimize the network structure. The results from the ablation study is summarized in **Table 3**. Due to memory limitation, we could not use more than one batch without additional down-sampling of the image dataset. We observed that using the drop-out without down-

sampling (batch size of one) showed better performance than batch normalization with down-sampling (batch size of two). Using the residual block and the element-wise sum array increased the Dice similarity coefficient. In addition, the dice coefficient was slightly improved by applying the drop-out for the proposed network.

Table 3. Ablation study between conventional 3D U-Net and proposed network

Model	Version	Dice Score
3D U-Net	+ Down-sampling (x2) + Batch normalization (batch size = 2)	0.848
	+ Drop-out ($P_{\text{drop}}^* = 0.3$)	0.862
U-Net	+ RB (Residual block), EWSA (Element-wise sum array)	0.885
	+ RB + EWSA + Drop-out ($P_{\text{drop}} = 0.3$)	0.890

* P_{drop} : Drop-out rate

Figures 3, and **4** show some cases in which the CNN outperformed the manual segmentation that was supported by the automatic inter-slice ROI interpolation function provided by the vendor’s software. Note that the errors are mainly associated to the time-consuming nature of the manual segmentation in which every frame was not segmented as a compromise. In **Figure 3**, the pelvis of the left kidney is wrongly included in the manual VOI (**Fig. 3A**) although the CNN did not yield such error (**Fig. 3B**).

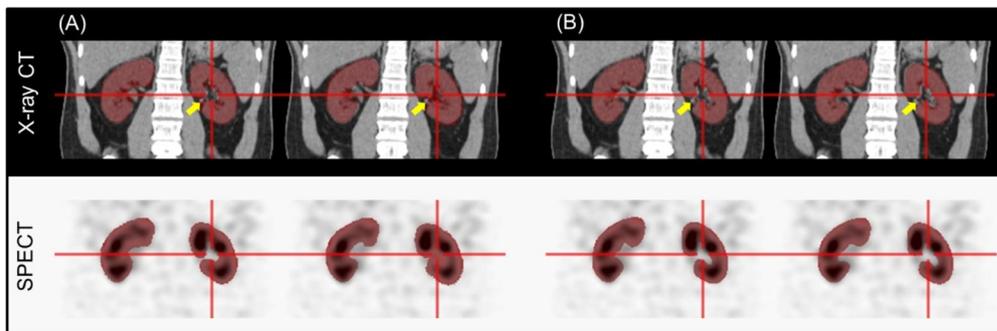


Figure 3. SPECT/CT images and renal parenchyma VOIs with incorrect ROI interpolation result (next slice as the second and fourth column) provided from the

vendor's software. (A) Manually segmented VOI. (B) Deep-learning-generated automatic VOI.

Figure 4 shows another case in which the CNN well excludes the multiple renal stones (yellow arrow in **Fig. 4B**), but the manual VOI failed to exclude multiple stones (**Fig. 4A**). The CNN well delineates the partial nephrectomy margin in the left kidney (red arrow in **Fig. 4D**). In addition, the accuracy of segmentation was better in the CNN outcome (yellow arrow in **Fig. 4C**).

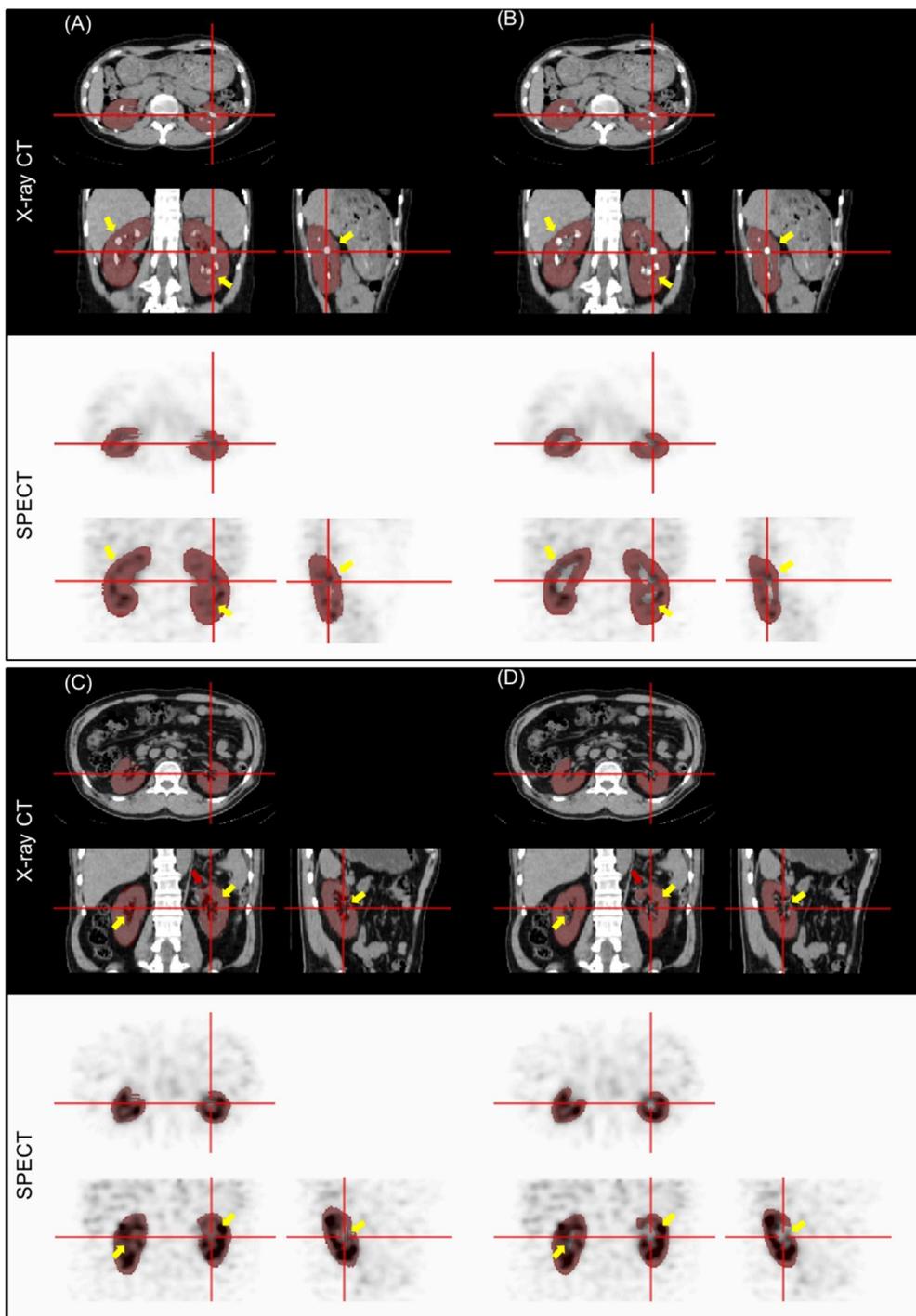


Figure 4. SPECT/CT images (multiple renal stones for (A), (B) and partial nephrectomy for (C), (D)) and renal parenchyma VOIs for a representative test dataset (A), (C) Manually segmented VOI. (B), (D) Deep-learning-generated automatic VOI.

3.2. GFR Estimation

The GFR values derived using manual and automatic segmentation methods were strongly correlated ($R^2 = 0.96$ in main experiment) for total kidneys (Table 2). Scattered and Bland–Altman plots between the measurement of GFR in total kidneys using manual and deep-learning-generated VOIs are shown in Figure 5A-C and D-F. (Total kidney in Fig. 5A and D, left kidney in Fig. 5B and E, right kidney in Fig. 5C and F)

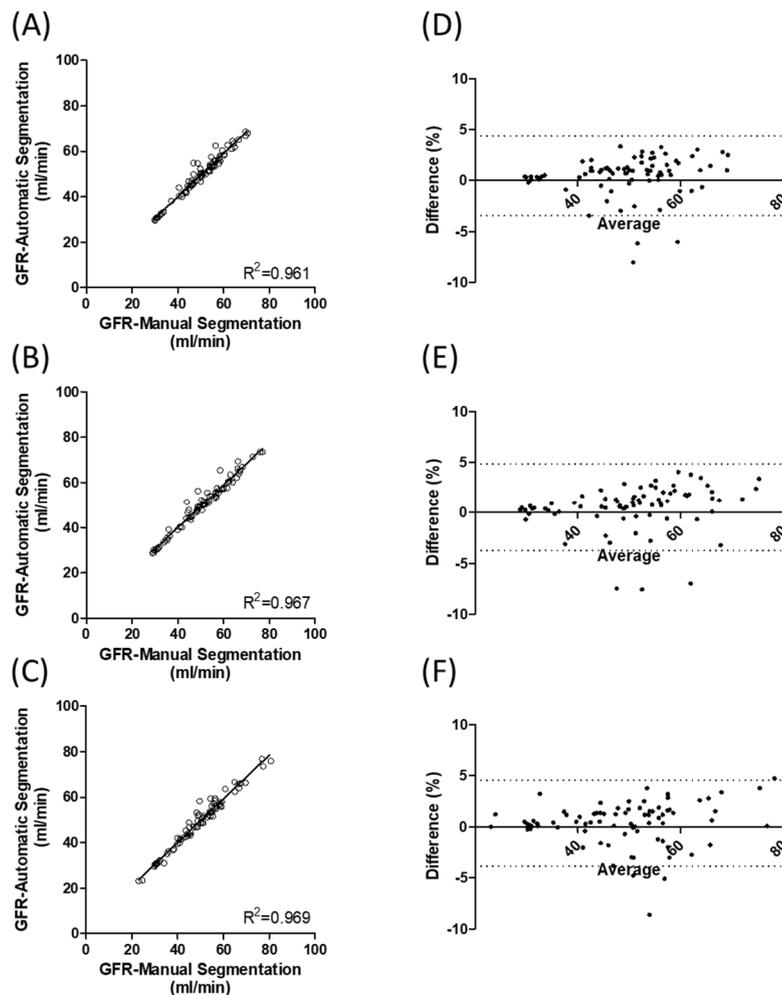


Figure 5. Scattered (A-C) and Bland–Altman (D-F) plots between measurement of left glomerular filtration rate (GFR) using manual and deep-learning-generated volumes of interest (VOIs). (A), (C) Total kidney. (B), (E) Left kidney. (C), (F) Right kidney.

Figure 6 shows the percentage difference (mean absolute percentage error) between the measurements of GFR obtained using manual and deep-learning-generated VOIs in all five-fold cross-validations. The absolute difference between the GFR values using manual (49.87 ± 10.08 ml/min) and automatic (49.41 ± 9.81 ml/min) methods was only $2.90 \pm 2.80\%$ (left kidney: $3.12 \pm 2.99\%$; right kidney: $3.13 \pm 2.80\%$) in the main experiment. The percentage differences obtained in the other cross-validations were $2.88 \pm 2.75\%$, $2.99 \pm 3.26\%$, $3.00 \pm 2.94\%$, and $2.68 \pm 2.70\%$, respectively. The results of cross-validations are summarized in **Table 2**. The correlation coefficient R^2 for the five sets ranged from 0.95 to 0.96.

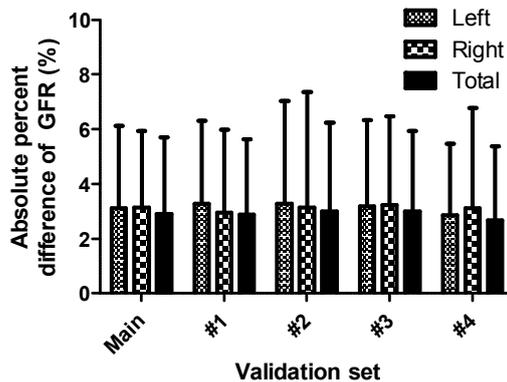


Figure 6. Absolute percentage difference between measurement of GFR using manual and deep-learning-generated VOIs: results of five-fold cross-validation.

3.3. Validation on Urolithiasis Patients

The CNN segmentation-based GFR (GFR_{CNN}) was applied for further clinical validation for urolithiasis patients. The manual-segmentation-based GFR by four human experts (GFR_{manual}) served as a reference. Individual kidney GFR and total GFR (sum of bilateral GFR with body surface area normalization) were investigated.

Figure 7A and **Table 4** show that GFR_{CNN} and GFR_{manual} were equivalent in terms of total GFR evaluation of urolithiasis and controls.

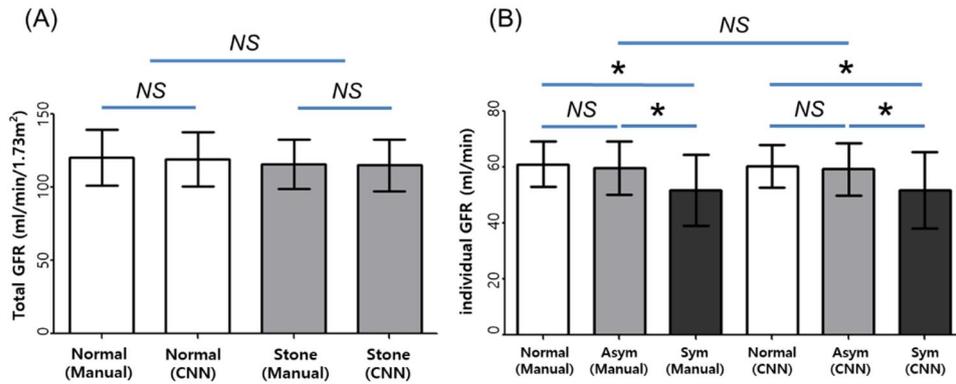


Figure 7. Total and Individual GFR comparison. (A) Total GFR comparison of manual and CNN-based automatic segmentation methods in kidney donor (normal) and urinary stone patients (stone). (B) Individual GFR comparison of normal, asymptomatic (Asym), or symptomatic (Sym) kidneys in manual and automatic segmentation methods. * $P < 0.001$.

Table 4. Total GFR (ml/min/1.73 m²) by manual and CNN-based segmentations in normal and urolithiasis patients (mean \pm SD)

	Normal ($n = 25$)	Stone ($n = 63$)	P -value
Manual	120.39 \pm 19.26	115.65 \pm 16.91	NS
CNN	119.25 \pm 18.35	115.02 \pm 17.71	NS
P -value	NS	NS	

NS, non-significant.

Total GFR_{CNN} in kidney donors (119.25 \pm 18.35 ml/min/1.73 m²) was not significantly different from GFR_{manual} (120.39 \pm 19.26 ml/min/1.73 m²; $P = 0.4432$, Wilcoxon test). Total GFR_{CNN} in urinary stone patients (115.02 \pm 17.71 ml/min/1.73 m²) was also not significantly different from GFR_{manual} (115.65 \pm 16.91 ml/min/1.73 m²; $P = 0.2387$, paired t -test). Meanwhile, total GFR in the normal and stone groups showed no significant difference in both manual ($P = 0.2582$, independent samples t -test) and CNN-based segmentations ($P = 0.5693$, Mann–Whitney test).

When it comes to the individual kidney GFR, GFR_{CNN} and GFR_{manual} were comparable with each other without significant difference. **Figure 7B** and **Table 5** show that individual GFR_{CNN} in normal kidneys (60.43 ± 7.66 ml/min) was not significantly different from GFR_{manual} (61.01 ± 8.10 ml/min; $P = 0.1725$, paired t -test).

Table 5. Individual GFR (ml/min) by manual and CNN-based segmentations in normal, symptomatic, and asymptomatic urinary stone kidneys (mean \pm SD)

	Normal ($n = 50$)	Asymptomatic ($n = 78$)	Symptomatic ($n = 48$)	P -value
Manual	61.01 ± 8.10	59.72 ± 9.46	51.84 ± 12.73	<0.001
CNN	60.43 ± 7.66	59.23 ± 9.25	51.76 ± 13.69	<0.001
P -value	<i>NS</i>	<i>NS</i>	<i>NS</i>	

* P -value less than 0.05/3 is considered significant in a comparison of manual versus CNN data (Bonferroni correction).

NS, non-significant.

In addition, individual GFR was not significantly different in asymptomatic kidneys (GFR_{CNN} : 59.23 ± 9.25 ml/min versus GFR_{manual} : 59.72 ± 9.46 ml/min; $P = 0.0361$, paired t -test) and in symptomatic kidneys (GFR_{CNN} : 51.76 ± 13.69 ml/min versus GFR_{manual} : 51.84 ± 12.73 ml/min).

Individual GFR in normal, asymptomatic, and symptomatic kidneys were significantly different in both manual ($P < 0.001$, ANOVA) and CNN-based segmentation methods ($P < 0.001$, ANOVA). *Post-hoc* analyses revealed that in both manual and CNN segmentations, symptomatic kidneys had significantly lower GFR compared with normal or asymptomatic kidneys ($P < 0.05$).

Finally, manual and automatic segmentation methods showed comparable performance in an evaluation of treatment response. **Figure 8** and **Table 6** present a typical case of a urinary stone patient before and after removal of ureter and renal

stones in a left kidney.

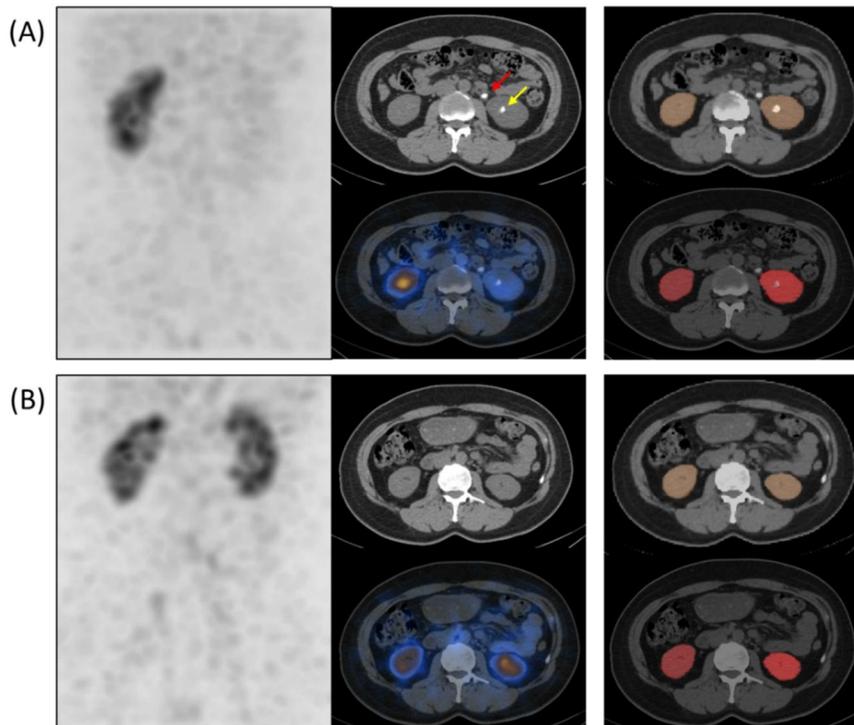


Figure 8. SPECT/CT images of a patient (A) before (red arrow indicates a ureter stone, yellow arrow indicates a renal stone) and (B) 4 months after removal of left ureter and renal stones. A projection image is presented in the first column, axial images of CT (top) and SPECT/CT fusion (bottom) are in the second, and segmentation results (automatic segmentation in the top and manual segmentation in the bottom) are shown in the third column.

Table 6. Change of individual kidney GFR before and after stone removal procedure in a urolithiasis patient

	Manual				CNN			
	%ID		GFR (g/ml)		%ID		GFR (g/ml)	
	Right	Left	Right	Left	Right	Left	Right	Left
Before surgery	4.13	1.54	60.84	37.15	4.22	1.64	61.64	38.05
After surgery	3.21	2.88	52.42	49.41	3.45	3.12	54.61	51.62

In serial projection images, ^{99m}Tc -DTPA uptake in left renal parenchyma is normalized after the procedure. Both manual and automatic segmentation methods showed marked improvement of %ID and individual GFR in the left kidney after

the removal of the stones.

Chapter 4. Discussion

In this study, I showed that the deep learning approach is highly accurate in renal parenchyma segmentation in CT images acquired in kidney SPECT/CT studies and is useful for automated measurement of GFR. The CNN outcomes yielded remarkably high Dice coefficient (0.89) with manual segmentation, leading to the strong correlations in %ID and GFR between the manual and automatic methods.

Automatically drawing VOIs only on renal parenchyma but excluding cysts and tumors is a challenging task because their CT intensities are very similar in non-contrast-enhanced CT images obtained in SPECT/CT studies. Although the proposed method performed the segmentation correctly in most cases as shown in **Figure 9**, there were several cases in which the segmentation was not accurate.

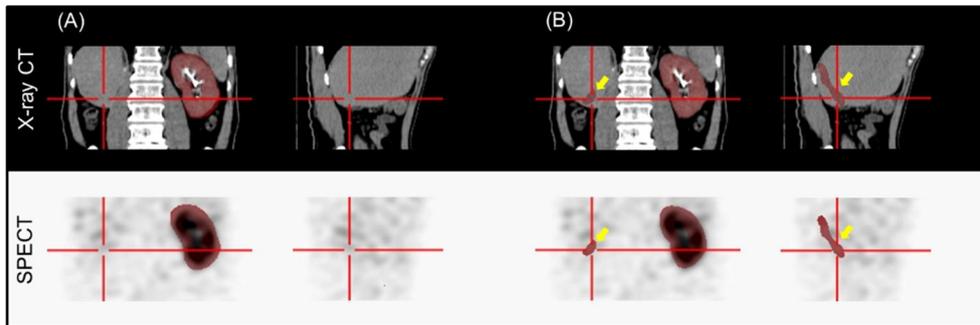


Figure 9. SPECT/CT images of a patient (A) before (red arrow indicates a ureter stone, yellow arrow indicates a renal stone) and (B) 4 months after removal of left ureter and renal stones. A projection image is presented in the first column, axial images of CT (top) and SPECT/CT fusion (bottom) are in the second, and segmentation results (automatic segmentation in the top and manual segmentation in the bottom) are shown in the third column.

Figure 10 is such a case in which a renal mass (yellow arrows) was incorrectly included although renal pelvis was well excluded (red arrows).

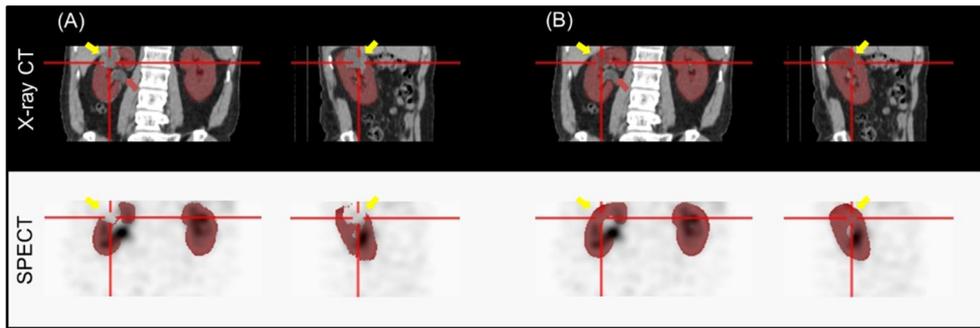


Figure 10. SPECT/CT images (renal mass and renal pelvis included) and renal parenchyma VOIs for a representative test dataset. (A) Manually segmented VOI. (B) Deep-learning-generated automatic VOI.

Because this patient (male, 164 cm, 58 kg) was relatively smaller than others, insufficient data for training deep CNN to properly handle such unusual cases would be the cause of inaccurate segmentation. In spite of such inaccuracy, the GFR error in this patient was only 2.48% because the radioactivity in the tumor was very low.

Because I trained the CNN to draw VOIs on the renal parenchyma of both kidneys, there was error in the patient with only a single kidney. In **Figure 11**, the CNN drew a long narrow VOI on the liver parenchyma (yellow arrow) of a patient who does not have a right kidney.

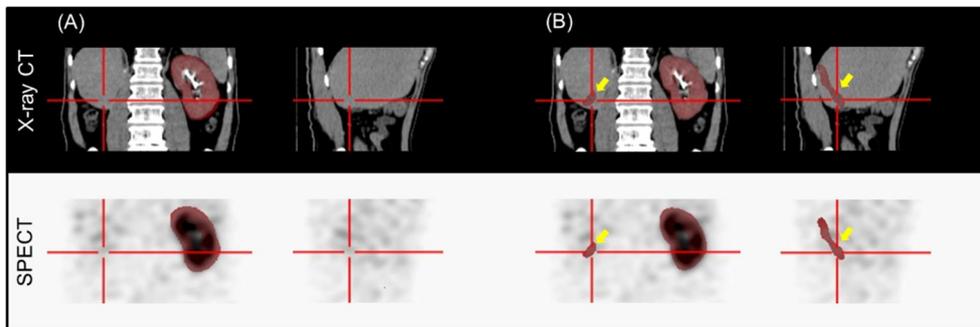


Figure 11. SPECT/CT images (single kidney) and renal parenchyma volumes of interest (VOIs) for a representative test dataset. (A) Manually segmented VOI. (B) Deep-learning-generated automatic VOI.

The CNN experienced only three single-kidney cases during the training among the training set with 272 patients. Additional datasets of with single-kidney patients will be necessary to overcome this limitation. In the cases shown in **Figures 3** and **4**, there were some segmentation errors in manual VOIs. By carefully inspecting the cause of these errors, I could reveal that the error in the manual VOI originated from the discontinuity in the perpendicular direction to the ROIs and subsequent inter-slice ROI interpolation. In two sequential slices shown in **Figure 3A**, the ROI on the left was manually drawn and that on the right was interpolated.

In the further validation of the proposed method for urolithiasis, automatic segmentation was comparable with manual segmentation in measuring GFR. There was no difference between manually driven GFR and CNN-driven GFR in all groups of patients (total GFR) and kidneys (individual GFR). In addition, in both automatic and manual segmentation methods, individual kidneys with symptomatic urinary stone had lower GFR compared with normal or kidneys with asymptomatic stones. I could presume that both segmentation methods work well to represent the functional deterioration by obstructing urinary stones and the subsequent improvement after stone removal procedures (31, 32). Further clinical validation of deep-learning-based segmentation is required to expand its use in more complicated cases such as multi-cystic dysplastic kidneys where manual segmentation is more laborious.

Chapter 5. Conclusion

The proposed deep learning approach to the 3D segmentation of kidney parenchyma in CT enables fast and accurate measurement of GFR. The combination of CT-based automatic segmentation by the deep learning approach and novel quantitative SPECT technology may pave the way for precision nuclear medicine regarding measurement of GFR.

Bibliography

- 1 Murray, A. W., Barnfield, M. C., Waller, M. L., Telford, T. & Peters, A. M. Assessment of glomerular filtration rate measurement with plasma sampling: a technical review. *J Nucl Med Technol* **41**, 67-75, doi:10.2967/jnmt.113.121004 (2013).
- 2 Matsushita, K., Mahmoodi, B. K., Woodward, M. & et al. Comparison of risk prediction using the ckd-epi equation and the mdrd study equation for estimated glomerular filtration rate. *JAMA* **307**, 1941-1951, doi:10.1001/jama.2012.3954 (2012).
- 3 Mulligan, J. S., Blue, P. W. & Hasbargen, J. A. Methods for measuring GFR with technetium-99m-DTPA: an analysis of several common methods. *J Nucl Med* **31**, 1211-1219 (1990).
- 4 Gates, G. F. Glomerular filtration rate: estimation from fractional renal accumulation of 99mTc-DTPA (stannous). *AJR Am J Roentgenol* **138**, 565-570, doi:10.2214/ajr.138.3.565 (1982).
- 5 Gates, G. F. Computation of glomerular filtration rate with Tc-99m DTPA: an in-house computer program. *J Nucl Med* **25**, 613-618 (1984).
- 6 Kang, Y. K. *et al.* Quantitative Single-Photon Emission Computed Tomography/Computed Tomography for Glomerular Filtration Rate Measurement. *Nucl Med Mol Imaging* **51**, 338-346 (2017).
- 7 Park, S. H., Lee, W. W., Oh, J. J. & Kim, S. E. Glomerular Filtration Rate Evaluation in Patients with Urinary Stone using Tc-99m DTPA Quantitative Single-Photon Emission Computed Tomography/Computed Tomography. *J Nucl Med* **57**, 1738 (2016).
- 8 Cha, K. H. *et al.* Urinary bladder segmentation in CT urography using deep learning convolutional neural network and level sets. *Med Phys* **43**, 1882-1896 (2016).
- 9 Chen, Y., Yu, W. & Pock, T. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5261-5269.
- 10 Han, Y. S., Yoo, J. J. & Ye, J. C. Deep Residual Learning for Compressed Sensing CT Reconstruction via Persistent Homology Analysis. *arXiv:1611.06391* (2016).
- 11 Hwang, D. H. *et al.* Improving the Accuracy of Simultaneously Reconstructed Activity and Attenuation Maps Using Deep Learning. *J Nucl Med* **59**, 1624-1629 (2018).
- 12 Kang, S. K. *et al.* Adaptive template generation for amyloid PET using a deep learning approach. *Hum Brain Mapp*, ; <https://doi.org/10.1002/hbm.24210> (2018).
- 13 Long, J., Shelhamer, E. & Darrell, T. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431-3440.

- 14 Park, J. Y. *et al.* Computed tomography super-resolution using deep convolutional neural network. *Phys Med Biol*, ; <https://doi.org/10.1088/1361-6560/aacdd1084> (2018).
- 15 Pinheiro, P. & Collobert, R. in *International conference on machine learning*. 82-90.
- 16 Teramoto, A., Fujita, H., Yamamuro, O. & Tamaki, T. Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Med Phys* **43**, 2821-2827 (2016).
- 17 Yoo, Y. S. On predicting epileptic seizures from intracranial electroencephalography. *Biomed Eng Lett* **7**, 1-5 (2017).
- 18 Im, C. H., Jun, S. C. & Sekihara, K. Recent advances in biomagnetism and its applications. *Biomed Eng Lett* **7**, 183-184, doi:10.1007/s13534-017-0042-3 (2017).
- 19 Sharma, K. *et al.* Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease. *Sci Rep* **7**, 2049 (2017).
- 20 Kayalibay, B., Jensen, G. & Smagt, P. v. d. CNN-based Segmentation of Medical Imaging Data. *arXiv:1701.03056* (2017).
- 21 Milletari, F., Navab, N. & Ahmadi, S.-A. in *3D Vision (3DV), Fourth International Conference on*. 565-571 (IEEE).
- 22 Kamnitsas, K. *et al.* Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* **36**, 61-78 (2017).
- 23 Dou, Q. *et al.* 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal* **41**, 40-54 (2017).
- 24 Suh, M. S., Lee, W. W., Kim, Y. K., Yun, P. Y. & Kim, S. E. Maximum Standardized Uptake Value of (99m)Tc Hydroxymethylene Diphosphonate SPECT/CT for the Evaluation of Temporomandibular Joint Disorder. *Radiology* **280**, 890-896, doi:10.1148/radiol.2016152294 (2016).
- 25 Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 424-432 (Springer).
- 26 He, K., Zhang, X., Ren, S. & Sun, J. in *European conference on computer vision*. 630-645 (Springer).
- 27 Mao, X., Shen, C. & Yang, Y.-B. in *Advances in neural information processing systems*. 2802-2810.
- 28 Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv:1603.04467* (2016).
- 29 An, H. J. *et al.* MRI-based attenuation correction for PET/MRI using multiphase level-set method. *J Nucl Med* **57**, 587-593 (2016).
- 30 Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization.

arXiv:1412.6980 (2014).

- 31 Keddis, M. T. & Rule, A. D. Nephrolithiasis and loss of kidney function. *Curr Opin Nephrol Hypertens* **22**, 390-396, doi:10.1097/MNH.0b013e32836214b9 (2013).
- 32 Mehmet, N. M. & Ender, O. Effect of urinary stone disease and its treatment on renal function. *World J Nephrol* **4**, 271-276, doi:10.5527/wjn.v4.i2.271 (2015).

Abstract in Korean

사구체여과율 (GFR)은 신장질환의 단계를 결정하고, 신장 기능의 단계를 측정하는데 가장 유용한 테스트 중 하나로 활용되고 있다. 이러한 GFR을 측정하는데 있어서 정량적인 단일광자 단층촬영/컴퓨터 단층촬영 (SPECT/CT)은 기존의 Planar scintigraphy에 비해 더욱 정확하고 신뢰할 수 있는 GFR의 측정이 가능하다. 그러나 SPECT/CT를 활용해 GFR을 측정하기 위해서는 CT 영상에서 신장 실질만을 관심 볼륨 (VOI)으로 그려야 하는 단점이 있다. 이러한 수동적인 작업은 숙련된 전문의가 20여분 정도 시간을 투자해야 하는 번거로운 작업이다. 본 연구의 목적은 딥러닝을 이용하여 CT 영상에서 신장 실질을 3D로 분할하고, GFR의 정량화를 자동화 하는 것이다.

정략적 ^{99m}Tc -DTPA SPECT/CT 촬영을 한 393명의 환자 데이터가 사용되었다. 정맥내 추적자 주사 후 2분뒤에 1분 SPECT 데이터가 수집된 이후 핵의학 전문의가 제조사의 프로그램을 이용해서 신장 실질의 VOI를 그렸다. 이렇게 그려서 분할된 신장 실질의 VOI와 CT 영상을 1대1로 학습할 수 있도록 수정된 3D U-Net이 딥러닝 네트워크 모델로 사용되었다. 정량적인 성능 평가를 위해서 수동으로 그린 결과와 딥러닝 출력 결과 간의 다이스 유사도 계수 (DSC)를 계산하였다. 또한 수동과 딥러닝의 분할된 신장 결과와 정량적인 SPECT영상을 이용해서 주입된 선량 백분율 (%ID)를 계산하였다. 이를 통해 GFR 측정 결과를

연고, 수동 분할과 딥러닝 분할 결과 간의 상관관계를 평가하였다. 또한 요로결석 환자 (63명)와 신장기증 예정 환자(25명)에서의 신장 분할의 임상적인 검증을 진행하였다. 총 173개의 신장을 방사선 비투과 결석 (Radio-opaque stone)의 존재 유무, 위치 (신장 혹은 요관), 크기 (10 mm 초과)에 따라 정상 (신장기증 예정 환자), 무증상 및 증상 (요로결석 환자) 신장 그룹으로 분류하였다. 각각 다른 환자 및 그룹의 신장에서 전체 및 개별 신장의 GFR을 통해 네트워크의 성능을 평가하였다.

제안하는 방법을 통해 CT 영상에서 자동으로 신장 분할을 할 수 있었고, 수동 분할과 비교하여 매우 높은 DSC (평균 0.89)를 얻을 수 있었다. 또한 GFR 측정 결과 역시 두 가지 분할 방법이 매우 밀접하게 상관관계가 있음 ($R^2 = 0.96$)을 확인할 수 있었다. 두 가지 분할 방법 간의 GFR 값의 절대 차이는 2.9%에 불과했다. 또한 요로결석 환자와 신장 기증 환자에서 두 가지 분할 방법을 통한 GFR결과가 통계적으로 유의함을 확인하였으며, 증상 신장이 무증상 혹은 정상 신장에 비해서 매우 낮은 GFR을 보였다.

본 연구에서 개발된 방법을 통해서 빠르고 정확한 GFR 측정을 할 수 있음을 확인하였다.

Keyword : 딥러닝, 신장 실질 분할, 사구체 여과율, 단일광자 단층촬영, 컴퓨터 단층촬영

Student Number : 2017-24168