Segmentation-Free Kidney Localization and Volume Estimation Using Aggregated Orthogonal Decision CNNs

Mohammad Arafat Hussain[1(✉)], Alborz Amir-Khalili[1], Ghassan Hamarneh[2], and Rafeef Abugharbieh[1]

[1] BiSICL, University of British Columbia, Vancouver, BC, Canada
{arafat,alborza,rafeef}@ece.ubc.ca
[2] Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC, Canada
hamarneh@sfu.ca

Abstract. Kidney volume is an important bio-marker in the clinical diagnosis of various renal diseases. For example, it plays an essential role in follow-up evaluation of kidney transplants. Most existing methods for volume estimation rely on kidney segmentation as a prerequisite step, which has various limitations such as initialization-sensitivity and computationally-expensive optimization. In this paper, we propose a hybrid localization-volume estimation deep learning approach capable of (i) localizing kidneys in abdominal CT images, and (ii) estimating renal volume without requiring segmentation. Our approach involves multiple levels of self-learning of image representation using convolutional neural layers, which we show better capture the rich and complex variability in kidney data, demonstrably outperforming hand-crafted feature representations. We validate our method on clinical data of 100 patients with a total of 200 kidney samples (left and right). Our results demonstrate a 55% increase in kidney boundary localization accuracy, and a 30% increase in volume estimation accuracy compared to recent state-of-the-art methods deploying regression-forest-based learning for the same tasks.

1 Introduction

Chronic kidney disease (CKD) refers to the reduced or absent functionality of kidneys for more than 3 months, which has been identified as a major risk factor for death worldwide [1]. It is reported that about 3 million adults in Canada [1] and about 26 million adults in the United States [2] live with CKD. Detection of CKD is difficult. Biochemical tests like the 'estimated glomerular filtration rate' and 'serum albumin-to-creatinine ratio' have been shown to be unreliable in detecting disease and tracking its progression [3]. However, CKD is most often associated with an abnormal change in kidney volume and thus, quantitative 'kidney volume' has emerged as a potential surrogate marker for renal function and has become useful for predicting and tracking the progression of CKD [4].

M. Descoteaux et al. (Eds.): MICCAI 2017, Part III, LNCS 10435, pp. 612–620, 2017.
DOI: 10.1007/978-3-319-66179-7_70

In addition, kidney volume has been used in evaluating the split renal function in kidney donors as well as in follow-up evaluation of kidney transplants [4].

Kidney volume measurement from 3D CT volumes in clinical environments typically requires a two step procedure: (i) localizing kidneys, and then (ii) estimating volumes of the localized kidneys. For years, kidneys were typically localized manually in clinical settings. To automate this process, Criminisi et al. [5,6] proposed regression-forest (RF)-based anatomy localization methods that predict the boundary wall locations of a tight region-of-interest (ROI) encompassing a particular organ. However, their reported boundary localization error for typical healthy kidneys (size $\sim 13 \times 7.5 \times 2.5\,\text{cm}^3$) was $\sim$16 mm, which could significantly affect the subsequent volume estimation, depending on the volume estimation process. Cuingnet et al. [7] fine tuned the method in [5] by using an additional RF, which improved the kidney localization accuracy by $\sim$60%. However, the authors mentioned (in [7]) that their method was designed and validated with only non-pathological kidneys. Recently, Lu et al. [8] proposed a right-kidney localization method using a cross-sectional fusion of convolutional neural networks (CNN) and fully convolutional networks (FCN) and reported a kidney centroid localization error of $\sim$8 mm. However, the robustness of this right-kidney-based deep learning model in localizing both kidneys is yet to be tested as the surrounding anatomy, and often locations, shapes and sizes of left and right kidneys are completely different and non-symmetric.

Subsequent to localization, kidney volumes are typically estimated using different segmentation methods, a strategy that has various limitations. For example, graph cuts and active contours/level sets-based methods are sensitive to the choice of parameters, prone to leaking through weak anatomical boundaries, and require considerable computations [4]. Recently, Cuingnet et al. [7] used a combination of RF and template deformation to segment kidneys, while Yang et al. [9] used multi-atlas image registration; these methods rely extensively on prior knowledge of kidney shapes. However, building a realistic model of kidney shape variability and deciding the balance between trusting the model *vs.* the data are non-trivial tasks. In addition, these prior-shape-based methods are likely to fail for pathological cases, e.g., presence of large exophytic tumors. To overcome these segmentation-based limitations and associated computational overhead, Hussain et al. [4] recently proposed a segmentation-free kidney volume estimation approach using a dual RF, which bypassed the segmentation step altogether. Although promising, their approach relied on manual localization of kidneys in abdominal CT. In addition, they used hand-engineered features that may be difficult to optimally design.

In this paper, we propose a hybrid deep learning approach that simultaneously addresses the combined challenges of (i) automatic kidney localization and (ii) segmentation-free volume estimation. Our method uses an effective deep CNN-based approach for tight kidney ROI localization, where orthogonal 2D slice-based probabilities of containing kidney cross-sections are aggregated into a voxel-based decision that ultimately predicts whether an interrogated voxel sits inside or outside of a kidney ROI. In addition, the second novel module in

our hybrid localization-segmentation approach is the direct estimation of kidney volume using a deep CNN that skips the segmentation procedure. To the best of our knowledge, our segmentation-free method is the first that uses deep CNN for kidney volume estimation. We estimate slice-based cross-sectional kidney areas followed by integration over these values across axial kidney span to produce the volume estimate. Our hybrid approach involves multi-level learning of image representations using convolutional layers, which demonstrably better capture the complex variability in the kidney data outperforming the hand-crafted feature representations used in [4,6,7].

2 Materials and Methods

2.1 Data

Our clinical dataset consisted of 100 patient scans accessed from Vancouver General Hospital (VGH) records with all the required ethics approvals in place. Of the 100 scans, 45 had involved the use of contrast agents. We accumulated a total of 200 kidney samples (both left and right kidneys) among which 140 samples (from 70 randomly chosen patients) were used for training and the rest for testing. Our dataset included 12 pathological kidney samples, and our training and test data contained 6 cases, each. The in-plane pixel size ranged from 0.58 to 0.98 mm and the slice thickness ranged from 1.5 to 3 mm. Ground truth kidney volumes were calculated from kidney delineations performed by expert radiologists. Both CNNs were implemented using *Caffe* [10] and were trained on a workstation with 2 Intel 3 GHz Xeon processors and an Nvidia GeForce GTX 460 GPU with 1 GB of VRAM. The base learning rate for both CNN training was set to 0.01 and was decreased by a factor of 0.1 to 0.0001 over 25000 iterations. For basic pre-processing of the data, we programmed an automatic routine for 'abdominal' CT that separates left and right kidneys. Since left and right kidneys always fall in the separate half volumes, the routine simply divided the abdominal CT volume medially along the left-right direction. Relying on the DICOM attributes (i.e., slice thickness and total number of axial slices), the routine also discarded few slices in the pelvic region from an image (where applicable). However, this step was optional and only carried out on slices beyond $\sim$52 cm (4x the typical kidney length) from the chest side of the image. Finally, our data pre-processing routine re-sized the medially separated CT volumes to generate fixed resolution cubic volumes (e.g., $256 \times 256 \times 256$ voxel in our case) using either interpolation or decimation, as needed. Augmentation of training samples was also done by flipping and rotating the 2D axial slices.

2.2 Kidney Localization

We use a deep CNN to predict the locations of six walls of the tight ROI boundary around a kidney by aggregating individual probabilities associated with three intersecting orthogonal (axial, coronal and sagittal) image slices (Fig. 1(b)). The

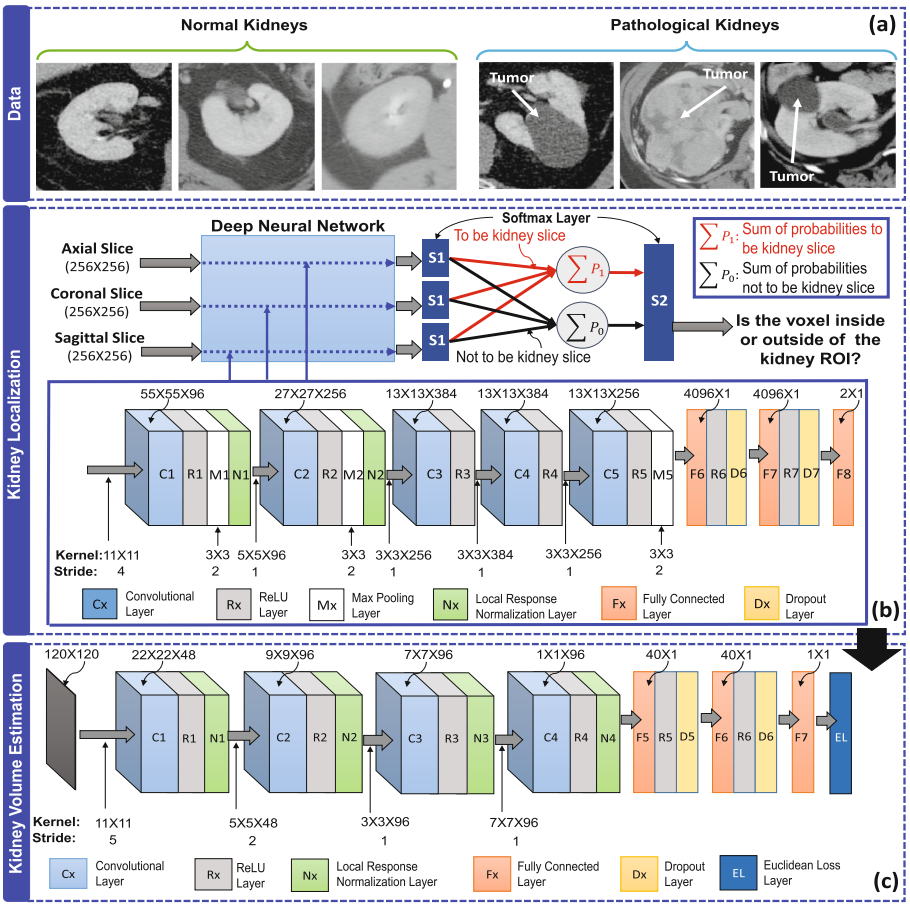

Fig. 1. Example kidney data from our patient pool demonstrating data variability present (ranging from normal to pathological), and our hybrid kidney localization-volume estimation approach. (a) Some CT snapshots showing variations in normal and pathological kidney shape and size, (b) orthogonal decision aggregated CNN for kidney localization, and (c) segmentation-free kidney volume estimation using deep CNN.

CNN has eleven layers excluding the input. It has five convolutional layers, three fully connected layers, two softmax layers and one additive layer. All but the last three layers contain trainable weights. The input is a 256×256 pixel image slice, either from the axial, coronal or sagittal directions, sampled from the initially generated local kidney-containing volumes. We train this single CNN (from layer 1 to layer 8) using a dataset containing a mix of equal numbers of 3D orthogonal image slices. Convolutional layers are typically used for sequentially learning the high-level non-linear spatial image features (e.g., object edges, intensity variations, orientations of objects etc.). Subsequent fully connected layers prepare

these features for optimal classification of the object (e.g., kidney cross-section) present in the image. In our case, five convolutional layers followed by three fully connected layers make reasonable decision on orthogonal image slices if they include kidney cross-sections or not. During testing, three different orthogonal image slices are parallel-fed to this CNN and the probabilities for each slice of being a kidney slice or not are acquired at the softmax layer, $S1$ (Fig. 1(b)). These individual probabilities from the three orthogonal image slices are added class-wise (i.e., containing kidney cross-section [class: 1] or not [class: 0]) in the additive layer, shown as $\sum P_1$ and $\sum P_0$ in Fig. 1(b). We then use a second softmax layer $S2$ with $\sum P_1$ and $\sum P_0$ as inputs. This layer decides whether the voxel where the three orthogonal input slices intersect is inside or outside the tight kidney ROI. The second softmax layer $S2$ is included to nullify any potential miss-classification by the first softmax layer $S1$. The voxels having probability ($\in[0,1]$) >0.5 at $S2$ are considered to be inside the kidney ROI. Finally, the locations of six boundary walls (two in each orthogonal direction) of this ROI are recorded from the maximum span of the distribution of these voxels (with probability >0.5) along three orthogonal directions.

2.3 Segmentation-Free Volume Estimation

In Sect. 2.2, we estimated the kidney encompassing tight ROI. Typically, kidney shape and appearance vary across patients (Fig. 1(a)). Training our CNN requires 2D image patches of consistent size. In addition, the patch size needs to be universal so that kidney cross-sections are always contained in it regardless of its size and shape. Therefore, to generate training data, we choose a patch size of 120×120 pixel, making sure that the cross-section of the initially estimated kidney ROI is at the centre of it. We also ensure that there is enough free space around a kidney cross-section as well as at the top and bottom in the axial direction. The ratio between the number of pixels fall inside a kidney cross-section to the total number of pixels in the image patch (120×120 pixel) is considered as the output variable (label) for that particular image patch.

We estimate the cross-sectional area of a kidney in each slice using a deep CNN shown in Fig. 1(c). The CNN performs regression and has seven layers excluding the input. It has four convolutional layers, three fully connected layers, and one Euclidean loss layer. We also use dropout layers along with the first two fully connected layers in order to avoid over-fitting. As mentioned earlier, the input is a 120×120 pixel image patch and the output is the ratio of kidney pixels to the total image size. The CNN is trained by minimizing the Euclidean loss between the desired and predicted values. Once the CNN model is trained, we deploy the model to predict the kidney area in a particular image patch. Finally, the volume of a particular kidney is estimated by integrating the predicted areas in all of its image patches in the axial direction.

3 Results

We provide results of our proposed kidney localization and volume estimation stages separately to enable direct comparisons with those obtained by recently reported kidney localization [5–8] and volume estimation methods [4,7,9,11,12]. Since the recently reported methods we use for comparison are mostly either RF-based or deep CNN-based, reproduce their results are impossible without access to the code and data on which they were trained. However, since the type of data these methods were validated is similar to ours in terms of resolution and imaging modality, we conservatively use their reported accuracy values for comparison, rather than using our own implementation of their models.

Table 1. Comparison of mean kidney ROI boundary localization error (mm) and mean kidney ROI centroid localization error (mm) in terms of Euclidean distance. Not reported values are shown with (-).

Methods	Boundary error (mm)		Centroid error (mm)	
	Left	Right	Left	Right
Cascaded RF (MICCAI'12) [7]	7.00 ± 10.0	7.00 ± 6.00	11.0 ± 18.0	10.0 ± 12.0
RF1 (MCV'10) [5]	17.3 ± 16.5	18.5 ± 18.0	-	-
RF2 (MedIA'13) [6]	13.6 ± 12.5	16.1 ± 15.5	-	-
CS-(CNN+FCN) (LABELS'16) [8]	-	-	-	7.80 ± 9.40
Proposed Method	$\mathbf{6.19 \pm 6.02}$	$\mathbf{5.86 \pm 6.40}$	$\mathbf{7.71 \pm 4.91}$	$\mathbf{7.56 \pm 4.10}$

In Table 1, we present kidney boundary and centroid localization performance comparisons of cascaded RF-based [7], single RF-based (RF1 [5] and RF2 [6]), cross-sectional (CS) fusion of CNN and FCN-based [8], and our proposed method. The cascaded RF method used the RF1 [5] for coarse localization of both left and right kidneys, then finetuned these locations using an additional RF per left/right kidney. Even then its centroid localization errors and boundary localization errors were higher than those of the proposed method. The RF2 [6] was an incremental work over the RF1 [5], and both use regression-forests for different anatomy localization. Both methods exhibited higher boundary localization errors than those of the cascaded RF and proposed methods, and did not report any centroid localization accuracy. The recently proposed CS-(CNN+FCN) [8] method reported significantly better kidney centroid localization performance with respect to the cascaded RF [7]. However, this method was only validated on right kidneys, and did not report the kidney boundary localization accuracy. As evident from the quantitative results, compared to all these recent methods, the proposed method demonstrates better performance in both kidney boundary and centroid localization by producing the lowest localization errors in all categories.

Table 2 shows quantitative comparative results of our direct volume estimation module (including the localization step) with those obtained by a

Table 2. Volume estimation accuracies compared to state-of-the-art competing methods. Not reported values are shown with (-).

Method types	Methods	Kidney samples	Mean volume error (%)	Mean dice index
Manual	Ellipsoid Fit (Urology'14) [11]	44	14.20 ± 13.56	-
Seg	RF+Template (MICCAI'12) [7]	358	-	0.752 ± 0.222 [a]
	Atlas-based (EMBS'14) [9]	22	-	0.952 ± 0.018
Seg-free	Single RF (MICCAI'14) [12]	44	36.14 ± 20.86	-
	Dual RF (MLMI'16) [4]	44	9.97 ± 8.69	-
	Proposed Method	**60**	$\mathbf{7.01 \pm 8.63}$	-

[a] Estimated from reported Dice quartile values (in [7]) using the method in [13].

manual ellipsoid fitting method, two segmentation-based methods, and two segmentation-free regression-forest-based methods. We use the mean volume errors by [11,12] reported in [4] for comparison. For the manual approach [11], we see in Table 2 that the estimated mean volume error for this approach is approximately 14% with high standard deviation. Then we consider two segmentation-based methods [7,9]. These methods reported their volume estimation accuracy in terms of Dice similarity coefficient (DSC), which does not relate linearly to the percentage of volume error. Since segmentation-free methods do not perform any voxel classification, DSC cannot be calculated for these methods. Therefore, it is difficult to directly compare DSC performance to percentage of volume estimation error. However, [7] used 2 RFs, an ellipsoid fitting and subsequent template deformation for kidney segmentation. Even then, authors in [7] admitted that this method did not correctly detect/segment about 20% of left and 20% of right kidneys (DSC <0.9, non-correctness criterion in [7]), and failed on about 10% left and 10% right kidneys (DSC <0.65, failing criterion in [7]). But our method successfully estimated volumes for all our kidney samples and achieved a mean volume error of 7%. Moreover, authors in [7] mentioned that the RF-based voxel classification was uncertain and the subsequent deformation step relies on the initial kidney shape. Due to this drawback, [7] is likely to fail on pathological kidneys. Crucially, authors in [7] did not include the truncated (tumor removal during partial nephrectomy) kidneys (16% of their data) in their evaluation. For similar reason, the multi-atlas image registration-based method [9] was evaluated only on 22 kidney samples out of 28, because 6 samples contained tumors. In addition, the test dataset size in [9] was very small (22 *vs.* 60 in our dataset). In contrast, our dataset includes 12 pathological kidney samples and our method does not fail for any kidney (pathological kidney volume estimation error is $6.96 \pm 8.67\%$) and thus, suggests it is less sensitive to kidney truncation or tumors. Finally, we consider two RF-based segmentation-free kidney volume estimation approaches [4,12]. For the single RF-based method [12], we see that the corresponding volume estimation error is worse than those of the dual RF and proposed methods as seen in Table 2. Using smaller 2D patches and the dual RF, [4] outperformed [12] but still shows higher volume estimation error

than the proposed method likely due to the use of non-optimal hand-engineered features. The last row of Table 2 reports the mean error of our proposed method around 7%, which is the lowest among all three segmentation-free methods.

4 Conclusions

We proposed a deep learning approach for hybrid localization-volume estimation for kidneys. This end-to-end method enabled a clinical approach for simultaneous (i) localization and (ii) segmentation-free volume estimation for kidneys from the raw abdominal CT data. We formulated an effective deep CNN-based method for kidney ROI localization, which aggregates 2D orthogonal slice-based kidney candidacy decisions. In addition, we formulated our volume estimation problem as a 2D image patch-based regression problem and were able to skip the often problematic segmentation step. Our deep CNNs better captured the rich and complex variability in kidney anatomy and outperformed the hand-engineered feature representations used in [4,6,7]. Our experimental results demonstrated a 55% increase in kidney boundary localization accuracy, and a 30% increase in volume estimation accuracy compared to those of recent literature [4,6].

Acknowledgement. We thank Dr. Timothy W. O'Connell and Dr. Mohammed F. Mohammed at VGH for providing the data and ground truth kidney tracing.

References

1. Arora, P., Vasa, P., Brenner, D., Iglar, K., McFarlane, P., Morrison, H., Badawi, A.: Prevalence estimates of CKD in Canada: results of a nationally representative survey. Canadian Med. Assoc. Jour. **185**(9), E417–E423 (2013)
2. Honeycutt, A.A., Segel, J.E., Zhuo, X., Hoerger, T.J., Imai, K., Williams, D.: Medical costs of CKD in the Medicare population. J. Am. Soc. Nephro. **24**(9), 1478–1483 (2013)
3. Connolly, J.O., Woolfson, R.G.: A critique of clinical guidelines for detection of individuals with chronic kidney disease. Neph. Clin. Pract. **111**(1), c69–c73 (2009)
4. Hussain, M.A., Hamarneh, G., O'Connell, T.W., Mohammed, M.F., Abugharbieh, R.: Segmentation-free estimation of kidney volumes in CT with dual regression forests. In: Wang, L., Adeli, E., Wang, Q., Shi, Y., Suk, H.-I. (eds.) MLMI 2016. LNCS, vol. 10019, pp. 156–163. Springer, Cham (2016). doi:10.1007/978-3-319-47157-0_19
5. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MCV 2010. LNCS, vol. 6533, pp. 106–117. Springer, Heidelberg (2011). doi:10.1007/978-3-642-18421-5_11
6. Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K.: Regression forests for efficient anatomy detection and localization in computed tomography scans. Med. Image Ana. **17**(8), 1293–1303 (2013)
7. Cuingnet, R., Prevost, R., Lesage, D., Cohen, L.D., Mory, B., Ardon, R.: Automatic detection and segmentation of kidneys in 3D CT images using random forests. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 66–74. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33454-2_9

8. Lu, X., Xu, D., Liu, D.: Robust 3D Organ Localization with Dual Learning Architectures and Fusion. In: International Workshop on Large-Scale Annotation of Biomedical Data Expert Label Synthesis, pp. 12–20 (2016)
9. Yang, G., Gu, J., Chen, Y., Liu, W., Tang, L., Shu, H., Toumoulin, C.: Automatic kidney segmentation in CT images based on multi-atlas image registration. In: IEEE Engineering in Medicine and Biology, pp. 5538–5541 (2014)
10. Jia, Y., E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: ACM International Conference on Multimodal, pp. 675–678 (2014)
11. Zakhari, N., Blew, B., Shabana, W.: Simplified method to measure renal volume: the best correction factor for the ellipsoid formula volume calculation in pretransplant CT live donor. Urology **83**(6), 1444.e15–1444.e19 (2014)
12. Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S.: Direct estimation of cardiac Bi-ventricular volumes with regression forests. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8674, pp. 586–593. Springer, Cham (2014). doi:10.1007/978-3-319-10470-6_73
13. Wan, X., Wang, W., Liu, J., Tong, T.: Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med. Res. Meth. **14**(1), 135 (2014)