Deep Learning Based Automatic Segmentation of Pathological Kidney in CT: Local vs. Global Image Context

Yefeng Zheng, David Liu, Bogdan Georgescu, Daguang Xu, and Dorin Comaniciu

Medical Imaging Technologies, Siemens Healthcare, Princeton, NJ, USA
yefeng.zheng@siemens.com

Abstract. Chronic kidney disease affects one of every ten adults in USA (over 20 million). Computed tomography (CT) is a widely used imaging modality for kidney disease diagnosis and quantification. However, automatic pathological kidney segmentation is still a challenging task due to large variations in contrast phase, scanning range, pathology, and position in the abdomen, etc. Methods based on global image context (e.g., atlas or regression based approaches) do not work well. In this work we propose to combine deep learning and marginal space learning (MSL), both using local context, for robust kidney detection and segmentation. Here, deep learning is exploited to roughly estimate the kidney center. Instead of performing a whole axial slice classification (i.e., whether it contains a kidney), we detect local image patches containing a kidney. The detected patches are aggregated to generate an estimate of the kidney center. Afterwards, we apply MSL to further refine the pose estimate by constraining the position search to a neighborhood around the initial center. The kidney is then segmented using a discriminative active shape model. The proposed method has been trained on 370 CT scans and tested on 78 unseen cases. It achieves a mean segmentation error of 2.6 mm and 1.7 mm for the left and right kidney, respectively. Furthermore, it eliminates all gross failures (i.e., segmentation is totally off) in a direct application of MSL.

1 Introduction

There are two bean-shaped kidneys in a normal person. Their main function is to extract waste from blood and release it from the body as urine. Chronic kidney disease (CKD) is the condition that a kidney does not function properly longer than a certain period of time (usually three months). In the most severe stage, the kidney completely stops working and the patient needs dialysis or a kidney transplant to survive. The incidence of CKD increases dramatically with age, especially for people older than 65 years. According to an estimate from the Center of Disease Control and Prevention, one in every ten American adults (over 20 million) have some level of CKD [1]. Computed tomography is a widely used imaging modality for kidney disease diagnosis and quantification. Different contrast phases are often used to diagnose different kidney diseases, including a native scan (no contrast at all to detect kidney stone), corticomedullary phase (much of contrast material still resides within the vascular system), nephrographic phase (contrast enters the collecting ducts), and excretory phase (contrast is excreted into the calices) [2].

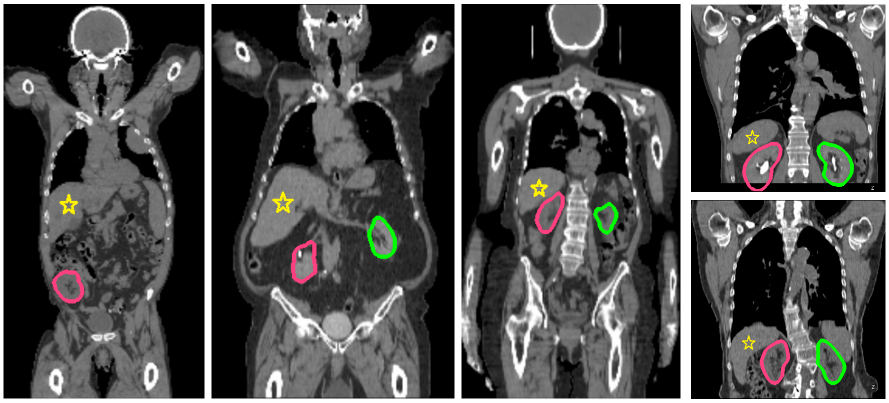


Fig. 1. Segmentation results of the left (green) and right (red) kidney. The relative position of the right kidney to the liver (yellow star) varies a lot as well as the scanning range. Note, the first patient has the left kidney surgically removed.

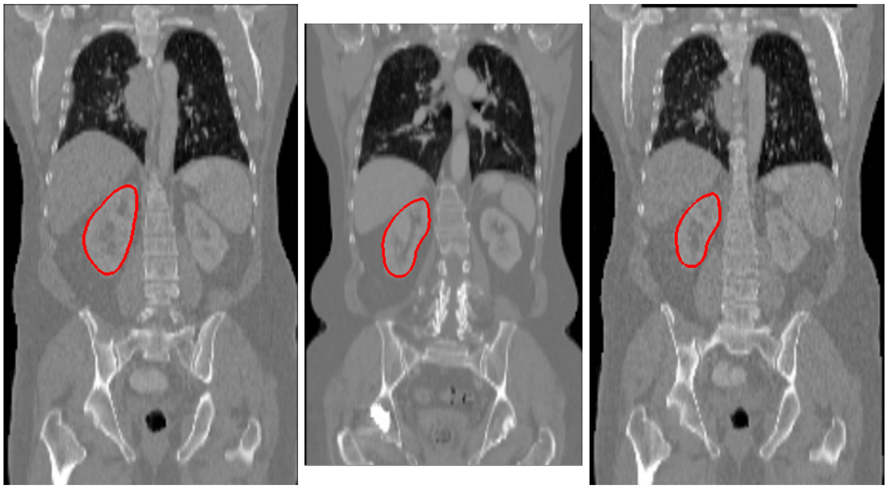
Various methods have been proposed to detect and segment an anatomical structure and many can be applied to kidney segmentation. Atlas based methods segment a kidney by transferring the label from an atlas to input data after volume registration [3]. However, volume registration is time consuming and several volume registrations are required in a multi-atlas approach to improve segmentation accuracy, but with increased computation time (taking several minutes to a few hours). Recently, regression based approaches [4–6] were proposed to efficiently estimate a rough position of an anatomical structure. An image patch cropped from anywhere inside a human body can be used to predict the center of a target organ by assuming a relatively stable positioning of organs. Regression is much more efficient than atlas registration and can estimate the rough position of an organ in a fraction of a second. Both atlas based and regression based approaches use global context to localize an organ. Alternatively, an organ can be detected via local classification in which we train a classifier that tells us if an image patch contains the target organ or not. Marginal space learning (MSL) [7, 8] is such an example, which efficiently prunes the pose parameter space to estimate the nine pose parameters (translation, orientation, and size) in less than a second. Recently, deep learning has been applied to kidney segmentation using the fully convolutional network (FCN) architecture [9]. However, it has only been validated on contrasted kidneys; therefore, its performance on more challenging dataset like ours is not clear.

Though existing methods may work well on normal kidneys, pathological kidney segmentation is still a challenging task. First, the relative position of a pathological kidney to surrounding organs varies. Normally, the right kidney lies slightly below the liver, while the left kidney lies below the spleen (as shown by the third and fourth patients in Fig. 1). However, in a patient with severe pathologies, a kidney may be pushed off by neighboring organs (e.g., liver, spleen, stomach, and colon) due to the excessive growth of tumor or previous surgery. For example, in our dataset, the relative position of the right kidney to the liver varies quite a lot, as shown in Fig. 1. The first patient has the right kidney lying at the bottom of the abdomen; while the kidney of the last patient resides at the top of the abdomen. Previous methods relying on global image context [3–6] cannot handle such large variation. Second, kidneys with severe pathologies exhibit extra variations in size, shape, and appearance. Last but not least, we want to develop a generic algorithm, which can handle all possible contrast phases in a renal computed tomography (CT) scan (as shown in Fig. 5). The native scan is especially difficult to segment due to the weak contrast between a kidney and the surrounding tissues.

Due to the large "floating" range of a pathological kidney inside the abdomen, a local classification based approach is more robust than an approach exploiting global context (e.g., global image registration [3] or using the liver or other organs to predict the kidney position [4–6]). For example, even though working well on detecting other large organs (e.g., liver and lungs), a regression based approach performs worse than MSL on kidney detection in a 3D magnetic resonance imaging (MRI) dataset [6]. Due to the challenges posed by our dataset, a direct application of MSL to CT kidney segmentation achieves a mixed result. It successfully detects and segments a kidney in 90-95% of cases. For the remaining cases, the segmentation may be completely off due to the failure in position detection, especially for patients with severe pathologies. We suspect that the limited success of MSL on pathological kidney detection is due to its use of hand-crafted features, which lack discriminative power to handle such large variations in our dataset.

In this work we propose to exploit deep learning for rough localization of the kidney. Deep learning can automatically build a hierarchical image feature representation, which has been shown in a lot of computer vision problems to outperform hand-crafted features. Recently, deep learning has been applied in many medical image analysis problems, including body region recognition [10], landmark detection [11], cell detection [12], lymph node detection [13], organ detection/segmentation [14, 15], cross-modality registration [16], and 2D/3D registration [17]. On all these applications, deep learning outperforms the state-of-the-art. In this work we apply deep learning to determine the abdomen range in a whole-body scan and then roughly localize the kidney inside the abdomen. Deep learning is especially data hungry, compared to other machine learning algorithms, to achieve good generalization capability. To mitigate the overfitting issue, we synthesize a lot of training data with realistic non-rigid deformations. After kidney localization, we apply MSL, but constrain the position search to a small range around the already detected kidney center. MSL also estimates the orientation and size of the kidney, thereby providing a quite good initial segmentation after aligning a pre-learned mean shape to the estimated pose. The segmentation is further refined using a discriminative active shape model (ASM) [7]. Please note, in this work, we treat the left and right kidney as different organs and train separate models to detect/segment them.

The remainder of the chapter is organized as follows. In Section 2 we present an approach to synthesize more training data with non-rigid deformation of the kidney. Abdomen range detection is presented in Section 3, which is used to constrain the search of the kidney. Kidney localization is described with detail in Section 4, followed by segmentation in Section 5. Quantitative experiments in Section 6 demonstrate the



(a) (b) (c)

Fig. 2. Synthesis of training images with non-rigid deformation of the right kidney. (a) Source volume with right kidney mesh overlaid. (b) Target volume. (c) Synthesized volume with intensity pattern from the source volume but the right kidney shape from the target volume.

robustness of the proposed method in segmenting pathological kidneys. This chapter concludes with Section 7.

2 Training Data Synthesis

To achieve good generalization on unseen data, deep learning needs a lot of training data; therefore, data augmentation is widely used to generate more training samples. Conventional data augmentation adds random translation, rotation, scaling, and intensity transformation, etc. In addition, we also add non-rigid deformation to cover variation in the kidney shape using an approach proposed in [18, 19]. Given two training volumes I_s and I_t with annotated kidney mesh S_s and S_t, respectively, we estimate the deformation field that warps S_s to S_t. The estimated deformation field is then used to warp all voxels in I_s to create a synthesized volume I_s^t. Here, we use the thin-plate spline (TPS) model [20] to represent the nonrigid deformation between the source and target volumes. The TPS interpolant $f(x, y, z)$ minimizes the bending energy of a thin plate

$$I_f = \int\int\int_{\mathcal{R}^3} \left(\frac{\partial^2 f}{\partial x^2}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2 + \left(\frac{\partial^2 f}{\partial z^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial z}\right)^2 + 2\left(\frac{\partial^2 f}{\partial y \partial z}\right)^2 dxdydz. \quad (1)$$

The interpolant $f(x, y, z)$ can be estimated analytically [20].

The kidneys in source and target volumes may be captured in different coordinate systems with different field-of-views. To avoid unnecessary coordinate changes, before estimating the TPS deformation field, we translate S_t so that, after translation, it has the same mass center as S_s.

The above TPS anchor points are concentrated on the kidney surface. To make the deformation field of background tissues smooth, we add the eight corners of the field-of-view of source volume I_s as additional TPS anchor points. Suppose the size of I_s is W, H, and D along three different dimensions, respectively. The following eight points are added as additional anchor points: $(0, 0, 0)$, $(W, 0, 0)$, $(0, H, 0)$, $(W, H, 0)$, $(0, 0, D)$, $(W, 0, D)$, $(0, H, D)$, and (W, H, D). These corner points will not change after deformation; therefore, the deformation field is strong around the kidney and gradually fades out towards the volume border.

Conceptually, the TPS deformation maps the source volume to the target volume. However, if we directly estimate this forward TPS warping and apply it to all voxels of the source volume, the resulting volume may have holes (i.e., voxels without any source voxels mapping to) unless we densely up-sample the source volume. Dense up-sampling the source volume increases the computation time when we perform forward TPS warping. In our implementation, we estimate the backward TPS warping, which warps the target volume to the source volume. For each voxel in the target volume, we use the backward TPS warping to find the corresponding position in the source volume. Normally, the corresponding source position is not on the imaging grid; so, linear interpolation is used to calculate the corresponding intensity in the source volume.

Fig. 2 shows an example of data synthesis for the right kidney. The source and target volumes are shown in Figs. 2a and b, respectively. The synthesized volume has an intensity pattern from the source volume but the right kidney shape from the target volume, as shown in Fig. 2c. The synthesized data is so visually realistic that it is difficult (if not impossible) to tell which one is real and which one is synthesized, by comparing Figs. 2a and c.

We have 370 training volumes. Since we can synthesize a new volume by taking an ordered pair of training data, we may synthesize up to $370 \times 369 = 136,530$ new training volumes. To reduce the training burden, we synthesize 2,000 new volumes by randomly picking training data pairs.

3 Abdomen Localization Using Deep Learning

Besides dedicated kidney scans, our dataset contains many whole-body CT scans, which are often acquired to diagnose cancer at multiple organs (e.g., to rule out metastasis). Depending on the exam indication, the scan range along the z axis (pointing from patient's toe to head) may be quite large and often varies a lot (as shown in Fig. 1). If we can constrain the MSL position detection to a limited range along the z axis, most detection failures of MSL can be eliminated. A kidney is bounded by the abdomen, though the position of a pathological kidney inside the abdomen varies as shown in Fig. 1. In this work we use a two-step approach to localize a kidney. We first determine the abdomen range and then detect a kidney inside the abdomen.

The abdomen has quite different image characteristics to other body regions (e.g., head, thorax, and legs); therefore, it can be detected reliably with an efficient classification scheme. We perform slice-wise classification by assigning a slice to one of three classes: above abdomen (head or thorax), abdomen, and legs. In our application, the lower limit of the abdomen stops at the top of the pubic symphysis (indicated by the

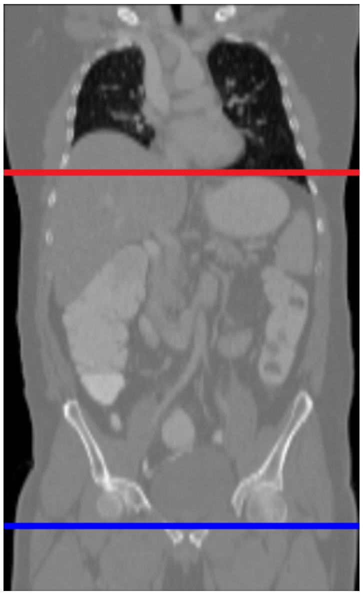


Fig. 3. Definition of abdomen range in a whole-body CT volume. The upper limit of the abdomen stops at the bottom of the heart; while, the lower limit of the abdomen stops at the top of the pubic symphysis.

blue line in Fig. 3), which joins the left and right pubic bones. With bony structures clearly visible in a CT volume, this landmark is easy to identify by a human being and, hopefully, also easy to detect automatically. The thorax and abdomen are separated by the diaphragm, which is a cursive structure as shown in Fig. 3. In this work we are not interested in the exact boundary. Instead, we use one axial slice to determine the upper limit of the abdomen. Here, we pick a slice at the bottom of the heart as the upper limit of the abdomen (the red line in Fig. 3).

A convolutional neural network (ConvNet) is trained to perform the slice-wise classification. To be specific, we use Caffe [21] to train a ConvNet with five layers of convolution and two fully connected layers (the “bvlc_reference_caffenet” model). A straightforward approach is to take a whole axial image as input to a classifier. However, the classifier may have difficulty in handling the variation of patient’s position inside a slice. Here, we first extract the body region (the white boxes in Fig. 4) by excluding the black margin. The input image is then resized to 227×227 pixels before feeding into the ConvNet.

Once the ConvNet is trained, we apply it to all slices in an input volume. For each slice we get a three-dimensional vector representing the classification confidence of each class (head-thorax, abdomen, and legs). To find the optimal range of the abdomen, we aggregate the classification scores as follows. Suppose we want to determine the bottom range of the abdomen (the boundary between the abdomen and legs); the input volume has n slices; and, the classification scores for the abdomen and leg classes are

$A[1, \ldots, n]$ and $L[1, \ldots, n]$, respectively. We search for the optimal slice index Ab_L such that

$$Ab_L = \underset{j}{\operatorname{argmax}} \sum_{i=1}^{j} (L[i] - A[i]) + \sum_{i=j+1}^{n} (A[i] - L[i]). \tag{2}$$

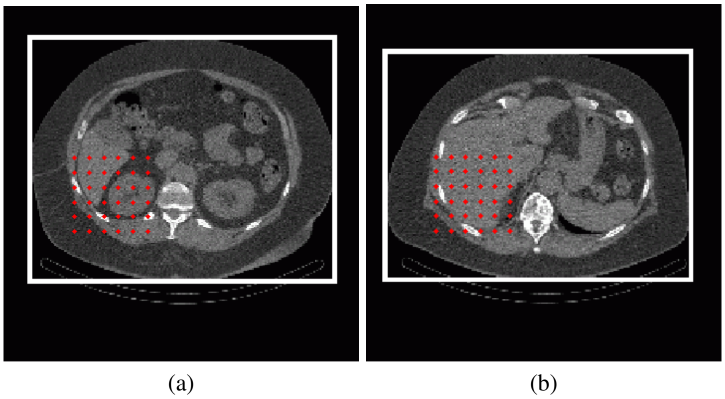
Below the abdomen/leg boundary, the leg class should have a higher score than the abdomen class. So, each individual item of the first term should be positive. Above the abdomen/leg boundary, the abdomen class should have a higher score; therefore, the second term should be positive. Equation (2) searches for an optimal slice index maximizing the separation between abdomen and legs. The upper range of the abdomen Ab_U is determined in a similar way.

$$Ab_U = \underset{j}{\operatorname{argmax}} \sum_{i=1}^{j} (A[i] - T[i]) + \sum_{i=j+1}^{n} (T[i] - A[i]). \tag{3}$$

Here, $T[1, \ldots, n]$ is the classification score of the thorax-head class. Aggregating the classification score of all slices, our approach is robust against noise in classification result on some slices.

Our approach is robust: the upper/lower range of the abdomen can be determined within an error of 1-2 slices without any gross failure. Using abdomen detection, we can quickly exclude up to 75% of slices from the following more through kidney localization procedure (which is far more time consuming). It accelerates the detection speed and, at the same time, reduces the kidney detection failures.

Previously, slice based classification is also used by Seifert et al. [22] to determine the body region. Our approach has several advantages compared to [22]. First, Seifert et al. formulated the task as a two-class classification problem, where the slice separating different body regions is taken as a positive sample and all other slices are treated as negative samples. Each training volume contributes only one positive training sample (maybe, a few after adding perturbation) and there are many more negative samples, which are often down-sampled to get a balanced training set. So, only a small number of slices are used for training. In our approach we formulate the task as a multi-class classification problem (i.e., thorax-head, abdomen, and legs). The distribution of different classes is more balanced; therefore, much more slices can be used for training. Second, using a two-class classification scheme, ideally, only the target slice should generate a high score and all other slices give a low score. If there is classification error on the target slice, the detection fails. In our approach, we aggregate the classification score of all slices to determine the boundary between body regions. Therefore, our approach potentially is more robust than [22]. Third, to separate the body into multiple regions, Seifert et al. train a binary classifier for each separating slice (in our case, two slices with one for the upper and the other for the lower range of the abdomen). During detection, all these classifiers need to be applied to all slices. Using a multi-class classification scheme, we apply a single classifier to each slice only once, which is more computationally efficient. Last but not least, [22] uses handcrafted Haar wavelet-like features for classification; while, we leverage the recent progress on deep learning, which can automatically learn more powerful hierarchical image features.



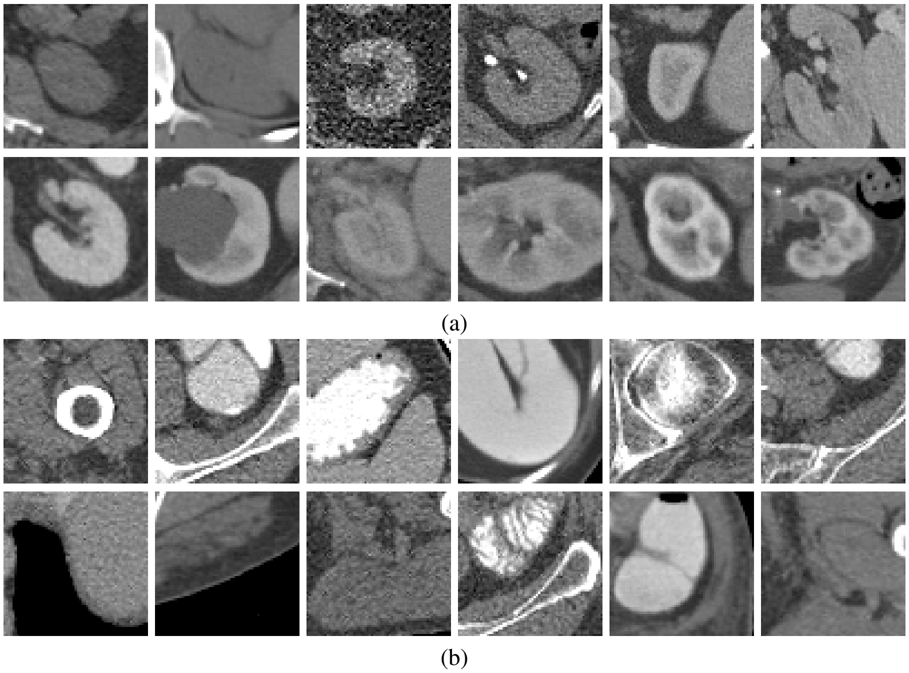
(a) (b)

Fig. 4. Patch centers (red dots) on a positive axial slice (a) and a negative slice (b) for right kidney localization. White boxes show the body region after excluding black margin.

4 Kidney Localization Using Deep Learning

Similar to abdomen detection, a classifier (e.g., ConvNet) can be trained to tell us if an axial image contains a kidney or not. However, this naive global-context approach generates a few gross failures. Assuming a kidney is next to the liver/spleen, a deep learning algorithm may use features from the liver/spleen to predict presence of a kidney in an axial image, which has a large input field-of-view covering kidney and surrounding organs. However, as shown in Fig. 1, the relative position of a pathological kidney to its surrounding organs is not stable. In this work we propose to crop a small image patch enclosing the kidney as input to a ConvNet. Since we do not know the exact position of the kidney, we need to test multiple patches. The red dots in Fig. 4 show the centers of cropped patches inside a predicted region-of-interest (ROI). During the training phase, we calculate the shift of the kidney center to the body region box center. The distribution of the shift helps us to define the ROI. As shown in Fig. 4, we crop $6 \times 6 = 36$ patches. Around each patch center, we crop an image of 85×85 mm^2, which is just enough to cover the largest kidney in our training set. For each positive slice, the patch with the smallest distance to the true kidney center is picked as a positive training sample. Afterwards, we randomly pick the same number of negative patches from slices without a kidney. Fig. 5 shows a few positive and negative training patches. Some negative patches are quite similar to positive patches (e.g., the last negative patch vs. the first two positive patches).

Similar to abdomen localization, we use Caffe [21] to train a ConvNet using the "bvlc_reference_caffenet" model. The standard input image size to this ConvNet model is 227×227 pixels. For patch based classification, we need to perform multiple classifications. To speed up the computation, we tried different input sizes and found that we could reduce the input to 65×65 pixels without deteriorating the accuracy. With a smaller input image size, we reduce the filter size of the first convolution layer from 11×11 to 7×7 and the stride from 4 to 2. All the other network parameters are kept the same.



(a)

(b)

Fig. 5. A few (a) positive and (b) negative training patches of the left kidney scanned with different contrast phases. Some negative patches are quite similar to positive patches (e.g., the last negative patch vs. the first two positive patches).

After classification of all axial slices, the ultimate goal is to have a single estimate of the kidney center. For whole slice or body region based classification, we only get one classification score for each slice. For a patch based approach, we have multiple patches and each has a classification score (a real positive value). We take the summation of scores of all patches that are classified positive as the final score of that slice. Negative patches do not contribute. A slice with more positive patches tends to have a higher score. Independent classification of each slice often generates noisy output (as shown in Fig. 6). We perform Gaussian smoothing with a kernel of 100 mm (the rough size of a kidney along the z axis). After smoothing, we pick the slice with the largest score as the kidney center along the z axis (Z_o). We then take positive patches on all slices within $[Z_o - 50, Z_o + 50]$ mm. The weighted average of the positive patch centers provides an estimate of the kidney center in the x and y axes.

5 Kidney Segmentation Based on MSL

After rough localization of a kidney, we use MSL to refine the position and further estimate its orientation and size. MSL is an efficient method for 3D anatomical structure detection and segmentation in various medical imaging modalities. The segmentation procedure is formulated as a two-stage learning problem: object pose estimation and

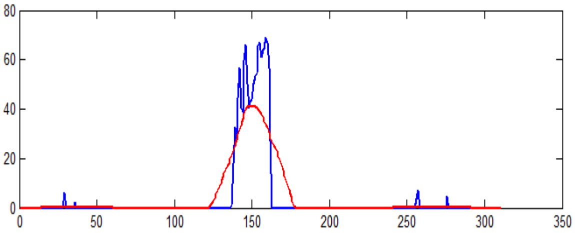


Fig. 6. Determining an axial slice containing the kidney center. The blue curve shows the aggregated classification score from multiple patches and the red curve shows the score after smoothing. The maximum peak on the red curve corresponds to the kidney center.

boundary delineation. To accurately localize a 3D object, nine pose parameters need to be estimated (three for translation, three for orientation, and three for anisotropic scaling). The object pose estimation is split into three steps: position estimation, position-orientation estimation, and position-orientation-size estimation. After each step only a small number of promising pose hypotheses are kept; therefore, the pose parameter space is pruned significantly to increase the detection efficiency. Since the kidney center has already been roughly determined using a ConvNet (Section 4), we constrain the MSL position search to a neighborhood around the initial center.

After the MSL based pose estimation, a mean shape is aligned to the estimated transformation to generate a rough estimate of the kidney shape. We then deform the shape to fit the kidney boundary using a machine learning based boundary detector within the ASM framework. Interested readers are referred to [7] for more details of the MSL based object detection and segmentation.

6 Experiments

We treat the left and right kidney as different organs and train separate models to detect/segment them. The systems are trained on 370 patients and tested on 78 patients (each patient contributes one CT scan). Our dataset is very diverse, containing various contrast phases and scanning ranges. Many patients have tumors inside the kidney or neighboring organs and some patients have previous abdominal surgery. The axial slice size is 512×512 pixels and the in-slice resolution varies from 0.5 mm to 1.5 mm, with a median resolution of 0.8 mm. The number of axial slices varies from 30 to 1239. The distance between neighboring slices varies from 0.5 mm to 7.0 mm with a median of 5.0 mm. On the test set, one patient has the left kidney surgically removed and three patients have the right kidney removed. These patients are ignored when we report the detection/segmentation error of the left and right kidney, respectively.

First, we evaluate the robustness of kidney localization using a ConvNet. There are far more negative training samples than the positives. We randomly subsample the negatives to generate a balanced training set with around 10,000 images for each class. We compare kidney localization errors of three input sizes: a whole slice, a body region, and a patch (85×85 mm^2). Since we only sample one training patch from each slice,

Table 1. Left kidney localization errors on 78 test cases with different input image context sizes.

	X		Y		Z	
	Mean	Max	Mean	Max	Mean	Max
Whole Slice	-	-	-	-	86.8	557.5
Body Region	-	-	-	-	14.5	112.5
Body Region (Multi)	-	-	-	-	12.0	45.7
Patch	**2.2**	**11.6**	**2.0**	**14.5**	**6.8**	**31.1**

Table 2. Right kidney localization errors on 78 test cases with different input image context sizes.

	X		Y		Z	
	Mean	Max	Mean	Max	Mean	Max
Whole Slice	-	-	-	-	113.6	631.5
Body Region	-	-	-	-	17.8	138.7
Body Region (Multi)	-	-	-	-	12.9	101.7
Patch	**3.1**	**46.9**	**3.0**	**17.5**	**7.9**	**56.7**

the number of training samples is the same for three scenarios, while the size of image context is different.

Tables 1 and 2 report the kidney center localization errors. The whole-slice based approach results in the worst performance with mean errors of 86.8 mm (the left kidney) and 113.6 mm (the right kidney) in determining the z-axis position of kidney center. Using the body region as input, we can significantly reduce the mean z-axis localization errors to 14.5 mm (the left kidney) and 17.8 mm (the right kidney). The patch-wise classification achieves the best result with mean z-axis localization errors of 6.8 mm (the left kidney) and 7.9 mm (the right kidney). In addition, it can accurately estimate the x and y position of the kidney center, with a mean error ranging from 2.0 to 3.1 mm. The larger mean errors in the z-axis are due to its much coarser resolution (a median resolution of 5.0 mm in z vs. 0.8 mm in x/y). For the left kidney localization, the maximum z-axis error is 31.1 mm. We checked this case and found the estimated position was still inside the kidney. (Please note, a typical kidney has a height of 100 mm along the z axis.) For the right kidney localization, there is one case that the estimated center is slightly outside the kidney. This error can be corrected later in the constrained position estimation by MSL.

For the patch based approach, we perform classification on 36 patches for each slice. One may suspect that its better performance comes from the aggregation of multiple classifications. To have a fair comparison, we also perform multiple classifications for the body region by shifting its center on a 6×6 grid (the same size as the patch grid). The results are reported as “Body Region (Multi)” in Tables 1 and 2. Aggregating multiple classifications improves the localization accuracy, but it is still worse than the proposed patch based approach. This experiment shows that local image context is more robust than global context in pathological kidney detection.

After rough localization of the kidney center using a ConvNet, we apply MSL to further estimate the nine pose parameters, followed by detailed boundary delineation using a discriminative ASM. Based on the error statistics in Tables 1 and 2, we constrain

Table 3. Kidney mesh segmentation errors on 78 test cases using marginal space learning with/without constrained position search range. The mesh errors are measured in millimeters, the smaller the better.

	Mean	Std	Median	Worst	Worst 10%
Left Kidney: Unconstrained	9.5	38.1	**1.3**	236.9	79.5
Left Kidney: Constrained	**2.6**	**4.2**	1.5	**24.7**	**11.6**
Right Kidney: Unconstrained	6.7	27.9	**1.4**	220.4	51.2
Right Kidney: Constrained	**1.7**	**1.2**	**1.4**	**6.8**	**4.6**

Table 4. Dice coefficient of kidney segmentation on 78 test cases using marginal space learning with/without constrained position search range. The Dice coefficient is in [0, 1], the larger the better.

	Mean	Std	Median	Worst	Worst 10%
Left Kidney: Unconstrained	0.86	0.24	**0.94**	0.00	0.21
Left Kidney: Constrained	**0.89**	**0.15**	0.93	**0.11**	**0.54**
Right Kidney: Unconstrained	0.88	0.19	0.93	0.00	0.46
Right Kidney: Constrained	**0.92**	**0.05**	**0.94**	**0.73**	**0.79**

the MSL position search to a neighborhood of $[-50, 50] \times [-20, 20] \times [-50, 50]$ mm^3 around the initial estimate. As shown in Table 3, we achieve a mean mesh segmentation error of 2.6 and 1.7 mm for the left and right kidney, respectively. The larger mean error of the left kidney is due to a case with a segmentation error of 24.7 mm. For comparison, without constraint, the mean segmentation errors of MSL are much larger due to some gross detection failures. The difference in the mean error of the worst 10% cases is more prominent: 11.6 mm vs. 79.5 mm for the left kidney and 4.6 mm vs. 51.2 mm for the right kidney. In Table 4, we also report the Dice coefficient. Unconstrained MSL has six gross failures (the segmentation has no overlap with ground truth resulting in a Dice coefficient of 0). All the failures are corrected by the proposed method.

It is hard to compare our errors with those reported in the literature due to the lack of a common test set. Lay et al. [6] reported that MSL outperformed their regression based approach on kidney detection in 3D MRI scans. Here, we achieve further improved robustness upon MSL. Cuingnet et al. [5] reported 6% of cases with Dice<0.65, while we have only three kidneys (2%) with Dice<0.65.

Our approach is fully automatic and takes about 3.3 s to detect a kidney: Kidney localization takes 2.8 s/volume on an NVIDIA GTX 980 GPU; The MSL detection/segmentation step takes 0.5 s on a computer with an Intel Xeon 6-core 2.6 GHz CPU and 32 GB memory (no use of GPU). Fig. 1 shows segmentation results on a few cases and more examples are shown in Fig. 7.

7 Conclusions

In this paper we proposed a robust fully automatic method for pathological kidney segmentation in CT scans. Deep learning is exploited to roughly estimate the kidney center, which is used to constrain the detection by MSL. We show that local image context

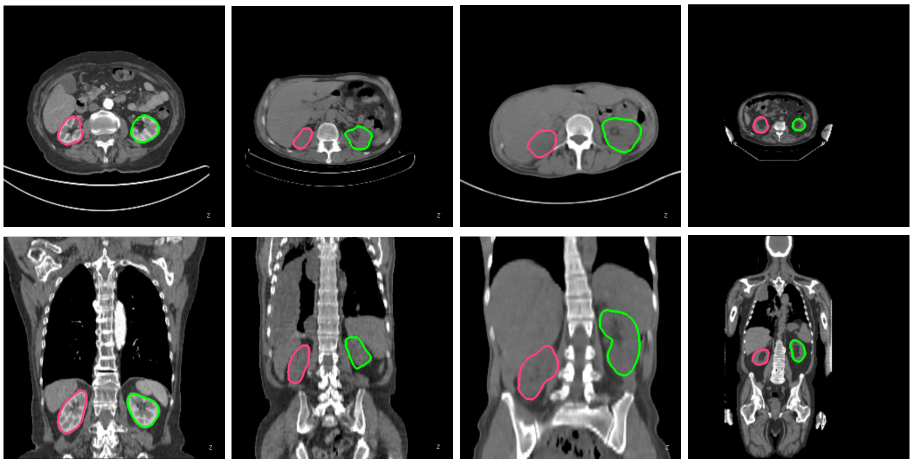


Fig. 7. A few examples of segmentation results of the left (green) and right (red) kidney. An axial view (top) and a coronal view (bottom) are shown for each example.

(small patches) is more robust than global context (whole slice or body region) in kidney detection and the proposed approach significantly reduces the number of gross failures. Our method works for renal CT data with different contrast phases, scanning ranges, and pathologies.

References

1. Center for Disease Control and Prevention: National chronic kidney disease fact sheet 2014 http://www.cdc.gov/diabetes/pubs/pdf/kidney_factsheet.pdf.
2. Yuh, B.I., Cohan, R.H.: Different phases of renal enhancement: Role in detecting and characterizing renal masses during helical CT. American Journal of Roentgenology **173**(3) (1999) 747–755
3. 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: Proc. Int'l Conf. IEEE Engineering in Medicine and Biology Society. (2014) 5538–5541
4. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention. (2011) 106–117
5. 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: Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention. (2012) 66–74
6. Lay, N., Birkbeck, N., Zhang, J., Zhou, S.K.: Rapid multi-organ segmentation using context integration and discriminative models. In: Proc. Information Processing in Medical Imaging. (2013) 450–462
7. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Medical Imaging **27**(11) (2008) 1668–1681
8. Zheng, Y., Comaniciu, D.: Marginal Space Learning for Medical Image Analysis – Efficient Detection and Segmentation of Anatomical Structures. Springer (2014)

9. Thong, W., Kadoury, S., Piche, N., Pal, C.J.: Convolutional networks for kidney segmentation in contrast-enhanced CT scans. In: Proc. of Workshop on Deep Learning in Medical Image Analysis. (2015) 1–8
10. Yan, Z., Zhan, Y., Peng, Z., Liao, S., Shinagawa, Y., Metaxas, D.N., Zhou, X.S.: Bodypart recognition using multi-stage deep learning. In: Proc. Information Processing in Medical Imaging. (2015) 449–461
11. Zheng, Y., Liu, D., Georgescu, B., Nguyen, H., Comaniciu, D.: 3D deep learning for efficient and robust landmark detection in volumetric data. In: Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention. (2015) 565–572
12. Liu, F., Yang, L.: A novel cell detection method using deep convolutional neural network and maximum-weight independent set. In: Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention. (2015) 349–357
13. Roth, H.R., Lu, L., Seff, A., Cherry, K.M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., Summers, R.M.: A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. In: Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention. (2014) 520–527
14. Carneiro, G., Nascimento, J.C., Freitas, A.: The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Trans. Image Processing. **21**(3) (2012) 968–982
15. Ghesu, F.C., Krubasik, E., Georgescu, B., Singh, V., Zheng, Y., Hornegger, J., Comaniciu, D.: Marginal space deep learning: Efficient architecture for volumetric image parsing. IEEE Trans. Medical Imaging (2016)
16. Cheng, X., Zhang, L., Zheng, Y.: Deep similarity learning for multimodal medical images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization (2016)
17. Miao, S., Wang, Z.J., Zheng, Y., Liao, R.: Real-time 2D/3D registration via CNN regression. In: Proc. IEEE Int'l Sym. Biomedical Imaging. (2016) 1–4
18. Zheng, Y., Doermann, D.: Handwriting matching and its application to handwriting synthesis. In: Int'l Conf. Document Analysis and Recognition. (2005) 1520–5263
19. Zheng, Y.: Cross-modality medical image detection and segmentation by transfer learning of shape priors. In: Proc. IEEE Int'l Sym. Biomedical Imaging. (2015) 424–427
20. Bookstein, F.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Machine Intell. **11**(6) (1989) 567–585
21. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. arXiv:1408.5093 (2014)
22. Seifert, S., Barbu, A., Zhou, K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciuc, D.: Hierarchical parsing and semantic navigation of full body CT data. In: Proc. of SPIE Medical Imaging. (2009) 1–8