Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets

Kenny H. Cha,[a] Lubomir Hadjiiski, Ravi K. Samala, Heang-Ping Chan, Elaine M. Caoili, and Richard H. Cohan
Department of Radiology, The University of Michigan, Ann Arbor, Michigan 48109-0904

(Received 9 February 2016; revised 4 March 2016; accepted for publication 7 March 2016; published 23 March 2016)

Purpose: The authors are developing a computerized system for bladder segmentation in CT urography (CTU) as a critical component for computer-aided detection of bladder cancer.
Methods: A deep-learning convolutional neural network (DL-CNN) was trained to distinguish between the inside and the outside of the bladder using 160 000 regions of interest (ROI) from CTU images. The trained DL-CNN was used to estimate the likelihood of an ROI being inside the bladder for ROIs centered at each voxel in a CTU case, resulting in a likelihood map. Thresholding and hole-filling were applied to the map to generate the initial contour for the bladder, which was then refined by 3D and 2D level sets. The segmentation performance was evaluated using 173 cases: 81 cases in the training set (42 lesions, 21 wall thickenings, and 18 normal bladders) and 92 cases in the test set (43 lesions, 36 wall thickenings, and 13 normal bladders). The computerized segmentation accuracy using the DL likelihood map was compared to that using a likelihood map generated by Haar features and a random forest classifier, and that using our previous conjoint level set analysis and segmentation system (CLASS) without using a likelihood map. All methods were evaluated relative to the 3D hand-segmented reference contours.
Results: With DL-CNN-based likelihood map and level sets, the average volume intersection ratio, average percent volume error, average absolute volume error, average minimum distance, and the Jaccard index for the test set were 81.9% ± 12.1%, 10.2% ± 16.2%, 14.0% ± 13.0%, 3.6 ± 2.0 mm, and 76.2% ± 11.8%, respectively. With the Haar-feature-based likelihood map and level sets, the corresponding values were 74.3% ± 12.7%, 13.0% ± 22.3%, 20.5% ± 15.7%, 5.7 ± 2.6 mm, and 66.7% ± 12.6%, respectively. With our previous CLASS with local contour refinement (LCR) method, the corresponding values were 78.0% ± 14.7%, 16.5% ± 16.8%, 18.2% ± 15.0%, 3.8 ± 2.3 mm, and 73.9% ± 13.5%, respectively.
Conclusions: The authors demonstrated that the DL-CNN can overcome the strong boundary between two regions that have large difference in gray levels and provides a seamless mask to guide level set segmentation, which has been a problem for many gradient-based segmentation methods. Compared to our previous CLASS with LCR method, which required two user inputs to initialize the segmentation, DL-CNN with level sets achieved better segmentation performance while using a single user input. Compared to the Haar-feature-based likelihood map, the DL-CNN-based likelihood map could guide the level sets to achieve better segmentation. The results demonstrate the feasibility of our new approach of using DL-CNN in combination with level sets for segmentation of the bladder.
 [http://dx.doi.org/10.1118/1.4944498]

Key words: computer-aided detection, deep-learning, segmentation, CT urography, bladder, level set

1. INTRODUCTION

Bladder cancer is the fourth most common cancer diagnosed in men. The American Cancer Society estimates that bladder cancer will cause 16 000 deaths (11 510 in men and 4490 in women) in the United States in 2015, with 74 000 new cases (56 320 in men and 17 680 in women) diagnosed, and early detection and treatment of bladder cancer increase patient survivability.[1]

Multidetector row CT (MDCT) urography is the imaging modality of choice for tracking urinary track abnormalities, as a single exam can be used to evaluate the kidneys, intrarenal collecting systems, and ureters.[2–6] Interpretation of a CT urography (CTU) study, however, requires extensive time. On average, 300 slices are generated for each CTU scan (range: 200–600 slices), and the radiologist interpreting the study has to visually determine if lesions are present within the urinary tracts. The possibility that multiple lesions may be present requires that the radiologist pays close attention throughout the entire urinary tract while frequently adjusting the displayed images to better visualize possible lesions. In addition, many different urinary anomalies may be found in a single CTU study. The radiologist has to identify and determine how likely each anomaly is an urothelial neoplasm. The challenges of analyzing a CTU study lead to a substantial variability among radiologists in detection of bladder cancer, with reported

sensitivities ranging from 59% to 92%.[7,8] Due to the workload of interpreting CTU studies, the chance for a radiologist to miss a subtle lesion may not be negligible, thus any technique that may help radiologists identify urothelial neoplasms within the urinary tract may be useful. Computer-aided detection (CAD) used as an adjunct may reduce the chance of oversight by the radiologists. We are developing a CAD system to detect bladder cancer in CTU, and bladder segmentation is a crucial step for such CAD systems. The segmented bladder defines the search region for the subsequent steps to detect lesion candidates. Thus any lesions excluded from the segmented bladder region will be missed during the detection step. On the other hand, nonbladder structures included in the segmented region will increase the possibility of false positive objects being detected. Therefore, accurate bladder segmentation that isolates the bladder from the surrounding anatomical structure is a critical component of a bladder cancer CAD system.[9]

Other researchers have attempted to segment the bladder on various imaging modalities. Li *et al.*[10] and Duan *et al.*[11] segmented the bladder wall from magnetic resonance (MR) cytoscopy in six patients and analyzed it for suspected lesions. In a different study, Duan *et al.*[12] developed a segmentation method using an adaptive window-setting scheme to detect tumor surfaces in MR images of ten patients. Han *et al.*[13] segmented the bladder wall in T1-weighted MR images using an adaptive Markov random field model and coupled level-set information in six patients. These methods are developed for MR images, which differ from the modality used in our study. In addition, the methods presented in these studies have not been validated with a larger data set. Chai *et al.*[14] developed a semiautomatic bladder segmentation method for cone beam CT images using population data as prior knowledge, using eight patients for training and 22 patients for validation. Using the population data, however, may result in poor performance for cases that have large deviations from the training set. These studies have smaller data sets compared to the study being presented.

There are challenges to segment bladders in CTU. Bladders may be filled with intravenous (IV) contrast material that partially or fully opacifies the bladder. The boundaries between the bladder wall and the surrounding soft tissue have very low contrast such that they are often difficult to delineate. In addition, bladders may be imaged in a variety of shapes and sizes. To address these challenges, Hadjiiski *et al.*[15,16] developed preliminary bladder segmentation methods for CTU using active contour with 15 patients and level sets with 70 patients. Hadjiiski *et al.*[17] also developed a segmentation package specifically designed based on the characteristics of the bladder in CTU images, referred to as conjoint level set analysis and segmentation system (CLASS), that segments the contrast-enhanced and noncontrast regions of the bladder separately, using two input bounding boxes, and then joins the regions together. They qualitatively evaluated the segmentation performance of 81 bladders and performed quantitative evaluation of 30 bladders comparing the computer-segmented contours to hand-segmented reference contours and obtained promising results. The CLASS method was further developed by Cha *et al.*[18] to improve the segmentation accuracy. Model-guided refinement was used to propagate the contours of the contrast-enhanced region if the level set propagation stopped prematurely due to substantial nonuniformity of the contrast. An energy-driven wavefront propagation that used changes in energies, smoothness criteria of the contour, and a stop criterion determined by the previous slice contour was designed to further propagate the conjoint contours to the correct bladder boundary. The segmentation performance was evaluated using 81 training cases and 92 independent test cases.

Convolutional neural networks (CNNs) have been used previously to classify patterns in medical images for use with computer-aided detection and specifically for microcalcification detection in mammograms.[19–26] In these applications, the training sets were typically small, generally using less than 500 samples. As computational power grows, CNNs with very complex architectures that require training with massive data become practical. The deep-learning CNN (DL-CNN) using graphics processing units (GPU) has been shown to be able to classify natural images using a large training set. Krizhevski *et al.*[27,28] have shown that by using DL-CNN, they are able to achieve relatively low error rates and good classification accuracy on the ImageNet ILSVRC-2010 and ILSVRC-2012 data sets,[29] and the CIFAR-10 data set.[30]

In this study, we explored the application of the DL-CNN to bladder segmentation. The DL-CNN was trained to recognize the patterns inside and outside the bladder and generated a bladder likelihood map to guide the level set segmentation. For comparison, we also generated a bladder likelihood map by using Haar features,[31,32] to differentiate the bladder region from the surrounding structures as classified by a random forest classifier. To evaluate the effectiveness of the template-based approach, their performances were compared to our previous CLASS with local contour refinement (LCR) method.

The paper is organized as follows. First, the data set used in the study is described. Second, the method of generating the bladder likelihood map using DL-CNN is presented. Third, the level set segmentation method using the likelihood map is described. Fourth, the method of generating the likelihood map using Haar features is designed as a comparison to the DL-CNN approach. Finally, the segmentation results are presented and discussed.

2. MATERIALS AND METHODS

A DL-CNN was trained to distinguish between regions of interest (ROI) that are inside and outside of the bladder. The DL-CNN outputs the likelihood that an input ROI is inside the bladder, which is used to form the bladder likelihood map. The map is used to generate the initial contour for level-set-based bladder segmentation. A flowchart of the segmentation method is shown in Fig. 1.

2.A. Data set

In this study, a data set of 173 patients undergoing CTU who subsequently underwent cystoscopy and biopsy was utilized. The cases were collected retrospectively from the Abdominal

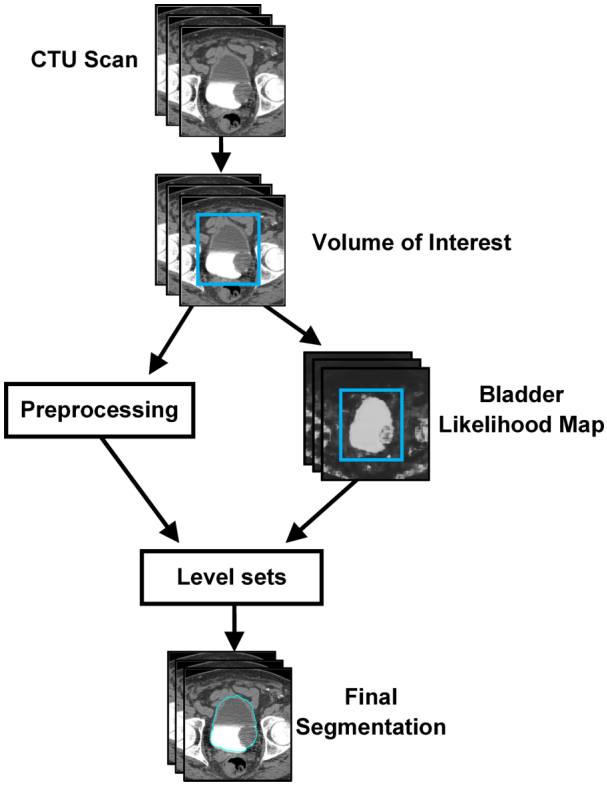


FIG. 1. Flowchart of the template-based segmentation method.

Imaging Division of the Department of Radiology at the University of Michigan with approval of the Institutional Review Board. We designated 81 of these cases as the training set, and the other 92 cases as the test set. The cases were assigned to the training or the test sets by balancing the difficulty of the cases between the two sets.

Of the 81 bladders in the training set, 42 contained focal mass-like lesions (40 malignant and 2 benign), 21 had wall thickening (16 malignant and 5 benign), and 18 were normal. Sixty-one bladders were partially filled with IV contrast material, 8 were completely filled with contrast material, and 12 had no visible contrast material. Of the 92 bladders in the test set, 43 contained focal mass-like lesions (42 malignant and 1 benign), 36 had wall thickening (23 malignant and 13 benign), and 13 were normal. Eighty-five bladders were partially filled with IV contrast material, four were completely filled with contrast material, and three had no visible contrast material. The bladder conspicuity was medium to high in both sets.

The CTU scans used in this study were acquired with GE Healthcare LightSpeed MDCT scanners. Excretory phase images, obtained 12 min after the initiation of the first bolus of a split-bolus IV contrast injection and 2 min after the initiation of the second bolus of 175 ml of nonionic contrast material at a concentration of 300 mg iodine/ml, were utilized. The images used were acquired using 120 kVp and 120–280 mA and reconstructed at a slice interval of 1.25 or 0.625 mm. Since patients were not turned prior to image acquisition, dependently layering IV contrast material that had been excreted into the renal collecting systems partially filled the bladder on the CTU images.

3D hand-segmented contours for all 173 cases were obtained as reference standard (RS1) in this study. An experienced radiologist provided manual outlines on the CT slices for all cases using a graphical user interface. The bladder was outlined on every 2D CT slice on which the bladder was visible, resulting in a 3D surface contour. There were a total of 16 197 slices for the 173 bladders. A subset of cases which contains lesions (41 training set cases, 50 test set cases, and a total of 8420 slices) were outlined by a different reader experienced in bladder segmentation to provide a second reference standard (RS2). The two sets of independent manual outlines allowed us to study the interobserver variability and to evaluate the difference in the computer segmentation performance relative to the two sets of hand-outlines.

2.B. Bladder likelihood map generation using DL-CNN

We applied the DL-CNN developed by Krizhevski *et al.* called CUDA-CONVNET (Refs. 27 and 28) to the classification of ROIs on 2D slices as being inside or outside of the bladder. The neural network is trained using labeled ROIs extracted from the CTU slices in the training cases. Each of the extracted ROIs is input into the DL-CNN, which outputs the likelihood of the ROI to be inside the bladder. To use the trained DL-CNN to generate a bladder likelihood map, it is applied to ROIs centered at each pixel on an axial slice in a CTU scan that contains the bladder and the likelihood value for the ROI is assigned to the center pixel. The resulting output over all pixels on the slice forms a bladder likelihood map, and the 2D maps over the consecutive CT slices constitute a 3D likelihood map.

2.B.1. DL-CNN components

Components of the DL-CNN are briefly described in the following. More information about this network can be found in the literature.[27,28]

2.B.1.a. Neurons. A DL-CNN neuron consists of two functional parts: (1) summation of the weighted inputs to the neuron and (2) application of an activation function to the sum. The activation function used in this DL-CNN is a nonsaturation nonlinear function, defined by the following equation:

$$f(x) = \max(0,x). \tag{1}$$

The output of a neuron generally is obtained by a sigmoid activation function; however, it was shown that networks trained with gradient descent can converge much faster when neurons with the activation function in Eq. (1) are used, which were named rectified linear units, following Nair *et al.*[28,33]

2.B.1.b. Convolution layer. In the convolution layer, the input ROI is convolved with the convolution kernels. The resulting values are collected into the corresponding neurons within the corresponding kernel maps in the convolution layer (Fig. 2). The output signals of these rectified linear unit neurons are generated using the activation function given by Eq. (1).

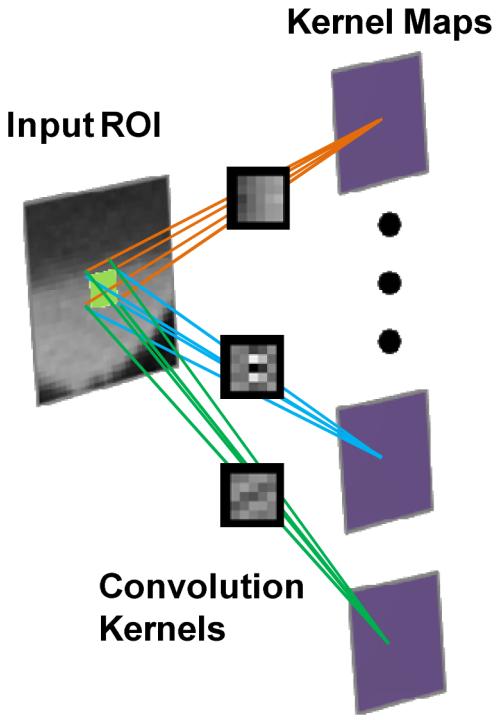


FIG. 2. Diagram of the convolution layer. An input ROI is convolved with multiple convolution kernels, and the resulting values are collected into corresponding neurons in the kernel maps.

2.B.1.c. Pooling layer. The pooling layers summarize the outputs of neighboring groups of neurons within the same kernel map. We compared two commonly used overlapping pooling for our application; one used the maximum values and the other used average values within 3×3 groups of pixels centered at the pooling unit, with the distance between pooling set to two pixels. It was found that using overlapping pooling was less prone to overtraining.[28]

2.B.1.d. Local response normalization layer. Using local normalization scheme aids in the generalization of the training. The activity of a neuron was normalized using the following equation:[28]

$$b^i_{x,y} = \frac{a^i_{x,y}}{\left(1+\frac{\tau}{N}\sum_{j=\max\left(0,i-\frac{N}{2}\right)}^{\min\left(n,i+\frac{N}{2}\right)}\left(a^j_{x,y}\right)^2\right)^{\varepsilon}}, \quad (2)$$

where $b^i_{x,y}$ is the response-normalized neuron activity, $a^i_{x,y}$ is the neuron activity computed by applying the kernel i at the coordinates (x,y), n is the number of kernel maps, and N, τ, and ε are constants. For our implementation of the DL-CNN, we used $N=9$, $\tau=0.001$, and $\varepsilon=0.75$, following the study by Krizhevsky *et al.*[28]

2.B.2. DL-CNN architecture

A block diagram of the network architecture used in this study is shown in Fig. 3. The network consists of five main layers: two convolution layers, two locally connected layers, and one fully connected layer. The locally connected layers

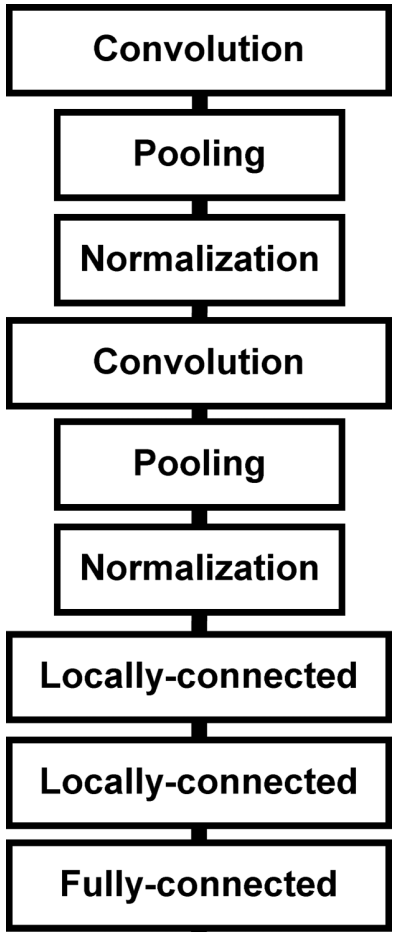


FIG. 3. Block diagram of the DL-CNN architecture used in this study.

perform the same operation as the convolution layer, except that instead of applying a single convolution kernel to every location of the input image to obtain a kernel map, different convolution kernels are applied at every location of the input image, and the resulting values are collected into the corresponding neurons within the corresponding kernel map. The fully connected layer uses every kernel map element multiplied by a weight as input. All of the inputs are summed, and the activation function [Eq. (1)] is applied to generate output values.

The first convolution layer filters the input images with 64 kernels of size 5×5. The output of the layer is pooled and normalized using the pooling and local response normalization and is input into the second convolution layer, which filters the output with additional 64 kernels of size 5×5. The first locally connected layer takes as input the pooled and normalized output of the second convolution layer and filters it with 64 kernels of size 3×3. The second locally connected layer has 32 kernels of size 3×3 connected to the normalized, pooled output of the first locally connected layer. The fully connected layer outputs two values. The outputs from the fully connected layer are input into a Softmax layer which computes the following function:

$$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}, \quad (3)$$

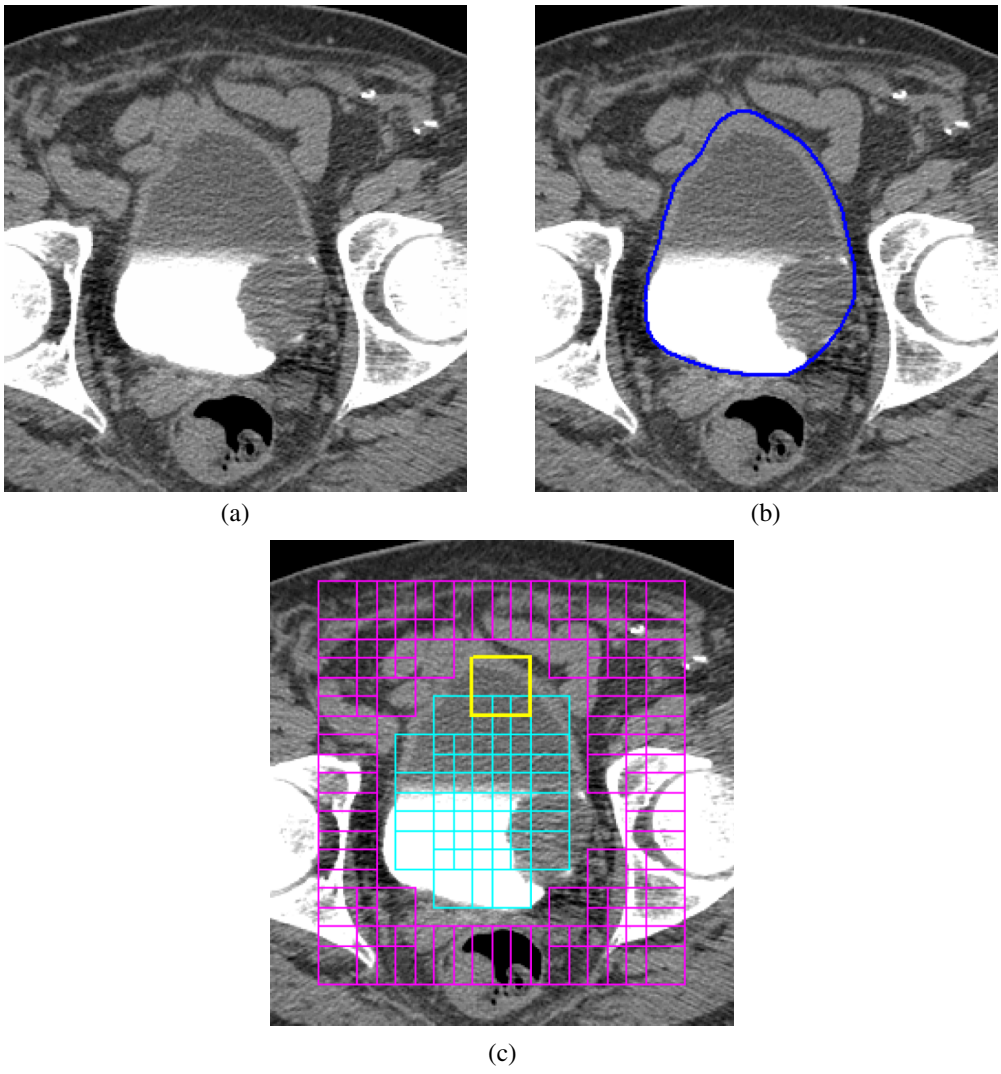
where x_i is each input value to the layer. The output of this layer ranges from 0 to 1, which can be interpreted as the likelihood of the input ROI being classified into one of the given categories.

2.B.3. DL-CNN training

The DL-CNN was trained using the cases in the training set. A cropped CTU slice of a bladder case is shown in Fig. 4(a). For each axial slice of the cases in the training set, ROIs of $N \times N$-pixels inside and outside the bladder were extracted using hand-outlines provided by an experienced radiologist [Fig. 4(b)]. Three ROI sizes $N = 16$, 32, 64 were studied but the size of 32×32 pixels was used in the following discussion. Each bladder ROI was labeled as being inside or outside of the bladder as follows. If over 90% of an ROI was within the hand-outlined bladder, the ROI was labeled as being inside the bladder. 90% was chosen to ensure that a sufficient number of ROIs is identified as being inside the bladder. If less than 5% of an ROI was within the hand-outlined bladder, the ROI was labeled as being outside the bladder to avoid most of the bladder and the bladder wall while including the background regions that surround the bladder. ROIs not labeled as being inside or outside of the bladder were excluded. Figure 4(c) shows examples of ROIs that were extracted from a slice.

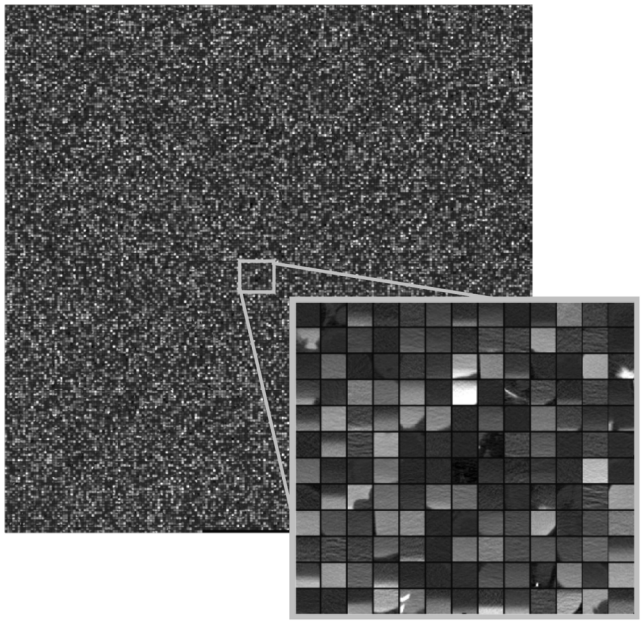
Approximately 160 000 ROIs were generated from the cases in the training set after balancing the number of ROIs that were inside and outside of the bladders. Figures 5(a) and 5(b) show examples of the ROIs inside and outside the bladder, respectively, used to train the DL-CNN.

The neural network was trained for 1500 iterations, but the DL-CNN trained for 1000 iterations was selected to generate the bladder likelihood maps. We observed that a network trained up to 1000 iterations had similar classification error rates to a network trained up to 1500 iterations. Classification error rate is defined as the ratio of the number of incorrectly identified ROIs to the total number of ROIs. Figure 6 shows the classification error rate of the DL-CNN training for the

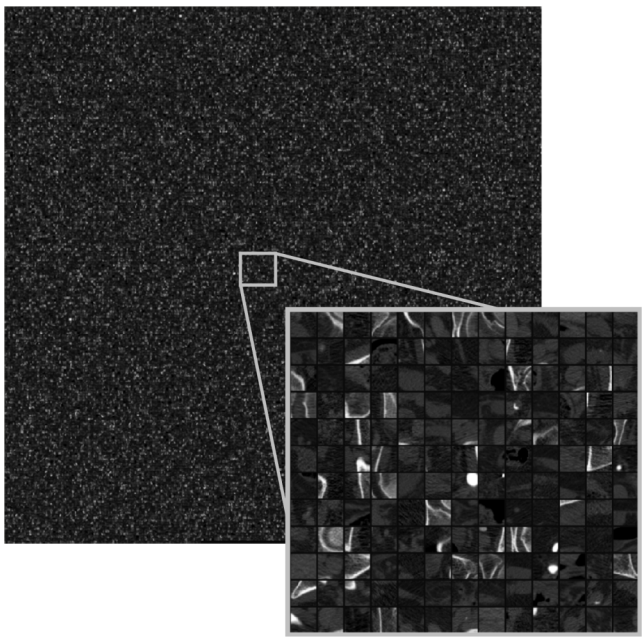


(a) (b) (c)

FIG. 4. Images of a CTU slice from a training case. (a) Cropped CTU slice centered at the bladder. (b) The CTU slice shown with radiologist's hand-outline of the bladder. (c) Example of ROIs that were extracted from the CTU slice to train the DL-CNN. The bright ROI at the top of the bladder shows the size of a 32×32-pixels ROI. The ROIs are partially overlapping. The darker ROIs are ones marked as outside of the bladder. The lighter ROIs are ones marked as inside of the bladder.



(a)



(b)

FIG. 5. Images of the 160 000 ROIs used to train the DL-CNN using the cases in the training set. Each ROI is 32×32 pixels. (a) ROIs that are labeled as inside the bladders. (b) ROIs that are labeled as outside the bladders. A small subset of the ROIs in each class is zoomed in to illustrate the content of typical ROIs.

entire training set as the number of iterations increased. In addition, we observed that bladder likelihood maps generated using DL-CNN trained for 1000 iterations were better or comparable to maps generated using network trained for 1500 iterations for representative cases of a range of difficulties in the training set; thus, 1000 iterations was used to generate the likelihood maps. Training the network using 160 000 ROIs and 1000 iterations took approximately 5.5 h using a Tesla C2075 GPU.

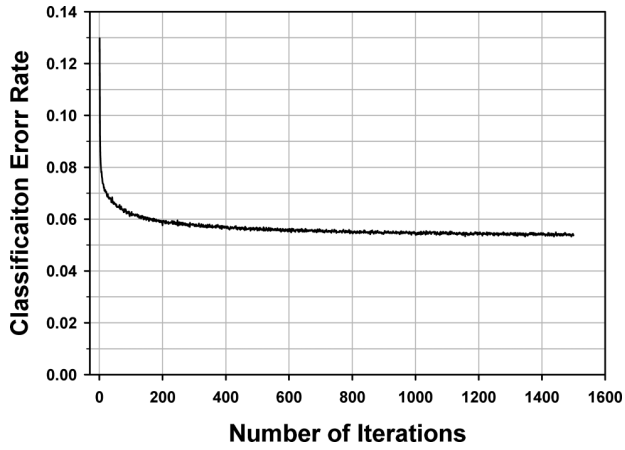


FIG. 6. Plot of the classification error rate of DL-CNN training for the entire training set as the number of iterations increases. The error rates at iterations 1000 and 1500 were very similar. The training results from iteration 1000 were used to generate the bladder likelihood maps.

2.B.4. Bladder likelihood map generation with DL-CNN

For every axial slice in a CTU scan that contains the bladder, a bladder likelihood map was generated. Our current segmentation system uses a single box or volume of interest (VOI) that approximately encloses the bladder as input. The bladder likelihood map is therefore generated within this VOI. The trained DL-CNN is applied to each voxel within the VOI. At each voxel, a 32×32-pixels ROI on the axial slice is extracted and input to the DL-CNN, which outputs the likelihood that the input ROI is inside the bladder. The likelihood score for the ROI is assigned to the center pixel of the ROI. The collection of voxelwise likelihood scores forms a bladder likelihood map. Figure 7 shows the bladder likelihood map of the CTU slice shown in Fig. 4.

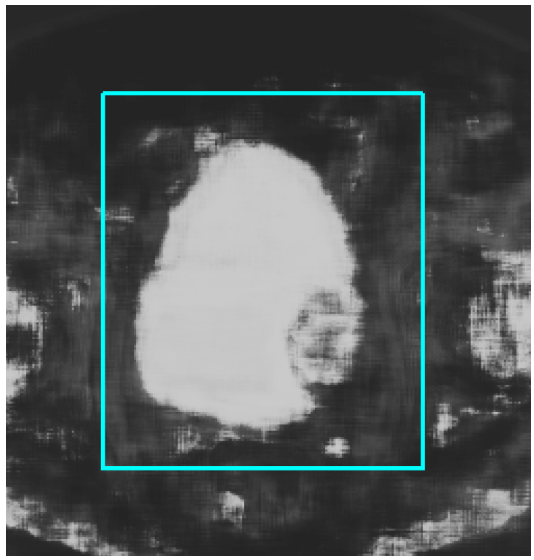


FIG. 7. Bladder likelihood map of the CTU slice shown in Fig. 4. High intensity represents high likelihood of the voxel being inside the bladder. In this example, for demonstration purposes, the bladder likelihood map was generated for an area larger than the VOI. The VOI is shown by the box around the bladder.

2.C. Bladder segmentation using DL-CNN bladder likelihood map

We are developing a software package that uses the DL-CNN bladder likelihood map and level sets to segment the bladder from the surrounding tissue. The system is initialized by the same VOI that encloses the bladder within which the bladder likelihood map is generated. The system consists of four stages: (1) preprocessing, (2) initial segmentation, (3) 3D level set segmentation, and (4) 2D level set segmentation.

In the first stage, preprocessing techniques are applied in 3D to the VOI. Smoothing, anisotropic diffusion, gradient filters, and the rank transform of the gradient magnitude are applied to the slices within the VOI to obtain a set of gradient magnitude images and a set of gradient vector images, which are used during level set propagation in the third stage.

In the second stage, the initial segmentation surface is generated using the DL-CNN bladder likelihood maps. First, a binary bladder mask, $\mathrm{DL}_{\mathrm{Mask}}$, is generated by applying the following criterion to every pixel on all slices of the bladder likelihood map:

$$\mathrm{DL}_{\mathrm{Mask}}(x,y)=\begin{cases}1, & \mathrm{DL}_{\mathrm{Score}}(x,y)\geq\theta\\ 0, & \mathrm{DL}_{\mathrm{Score}}(x,y)<\theta\end{cases}, \tag{4}$$

where $\mathrm{DL}_{\mathrm{Mask}}(x,y)$ is the pixel value on the bladder mask at the coordinates (x,y), $\mathrm{DL}_{\mathrm{Score}}(x,y)$ is the bladder likelihood score at the coordinates (x,y), and θ is the threshold imposed on the bladder likelihood score. The value for θ was determined by histogram analysis. A histogram of the DL-CNN likelihood score for the pixels inside and outside of the bladder within the VOIs in the training cases was generated (Fig. 8). We observed that the likelihood score of 0.85 provided a good separation of the two classes (e.g., inside the bladder and outside the bladder), with a large number of pixels correctly identified as being inside the bladder. Thresholding the likelihood maps at the score of 0.85 gave the best contour that did not leak to the outside of the bladder while closely approaching the hand segmentation for cases in the training set. For these reasons, 0.85 was chosen as the threshold.

Second, an ellipsoid whose minor and major axes are 1.5 of the width and height of the VOI, respectively, centered at the centroid of the bladder mask, is placed on the $\mathrm{DL}_{\mathrm{Mask}}$. The intersection of the bladder mask and the ellipsoid is labeled as the object region. The ellipsoid is used to prevent the object region from leaking into the organs above the bladder and the pelvic bone, as these structures can also obtain high likelihood scores from the DL-CNN. Finally, a morphological dilation filter with a spherical structuring element of two voxels in radius, 3D flood fill algorithm, and a morphological erosion filter with a spherical structuring element of two voxels in radius is applied to the object region to connect neighboring components and extract an initial segmentation surface, $\phi_0(x)$.

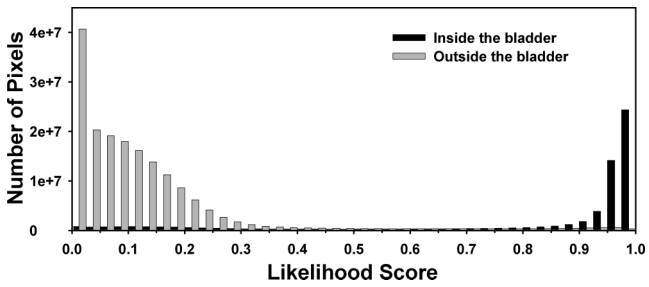


FIG. 8. Histogram of the DL-CNN likelihood score for the pixels in the training set. Higher likelihood score indicates that the pixel is more likely to be inside the bladder.

TABLE I. Parameters for the level sets.

Level set	α	β	γ	n
First	1	2	1	10
Second	1	0.6	q	150
Third	0	1.0	0	10
2D slices	4.0	0.2	0.5	100

In the third stage, the initial segmentation surface is propagated toward the bladder boundary using cascading level sets. Our chosen level set implementation evolves according to the equation,

$$\begin{cases}\dfrac{\partial}{\partial t}\Psi(x)=-\alpha A(x)\nabla\Psi(x)-\beta P(x)|\nabla\Psi(x)|+\gamma\kappa(x)|\nabla\Psi(x)|\\ \Psi(x,n=0)=\phi_0(x)\end{cases}, \tag{5}$$

where α, β, and γ are the coefficients for the advection, propagation, and curvature terms, respectively, $A(x)$ is a vector field image (assigning a vector to each voxel in the image) which drives the contour to move toward regions of high gradient, $P(x)$ is a scalar speed term between 0 and 1 causing the contour to expand at the local rate, and $\kappa(x)=\mathrm{div}(\nabla\Psi(x)/|\nabla\Psi(x)|)$ is the mean curvature of the level set at point x. The symbol ∇ denotes the gradient operator and div is the divergence operator.[34] $\phi_0(x)$ is the initial segmentation surface, and n is the number of iterations.

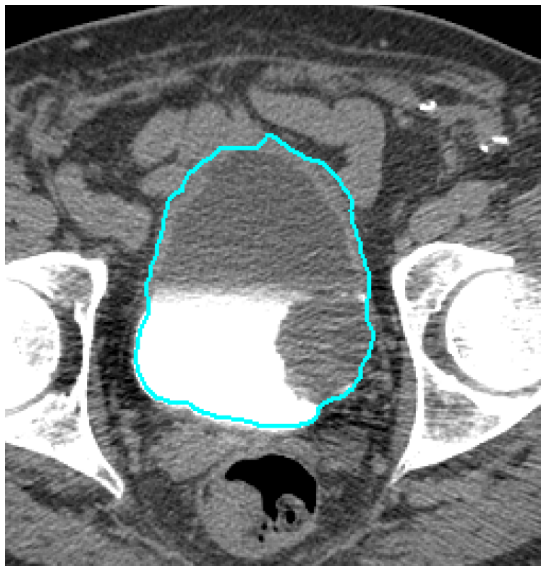


FIG. 9. Bladder segmentation of the CTU slice shown in Fig. 4 using the DL-CNN bladder likelihood map with level sets.

TABLE II. Number of features extracted for different Haar filter sizes and filter types as described by Viola *et al.* and Lienhart *et al.*

	8 × 8-pixels	16 × 16-pixels	16 × 18-pixels	18 × 16-pixels	16 × 32-pixels	32 × 16-pixels	32×32-pixels
Edge features	10	8	0	0	2	2	2
Line features	10	0	4	4	0	0	2
Four-rectangle features[a]	9	5	0	0	0	0	1

[a]A single filter of this four-rectangle feature filter consists of four smaller, equal-sized rectangles arranged in a checkerboard pattern.

Three 3D level sets with predefined sets of parameters are applied in series to the initial segmentation surface. The corresponding parameters of the three level sets are presented in Table I.

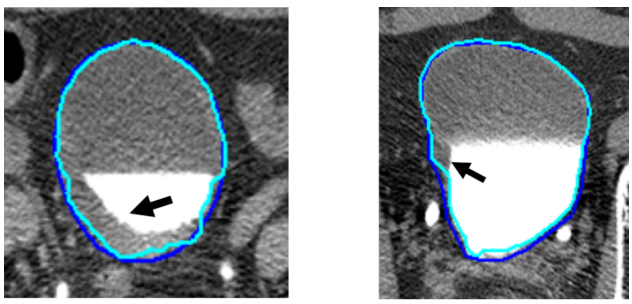
The first 3D level set slightly expands and smoothens the initial contour. The second 3D level set brings the contour toward the sharp edges but also expands it slightly in regions of low gradient. The parameter "q" in Table I is defined to be a linear function $\sigma M + \phi$ of the 2D diagonal distance M of the VOI box in millimeters (mm), where $\sigma = 0.06$, $\phi = -0.11$ as shown previously.[34] The third 3D level set further draws the contour toward sharp edges. As a final step, a 2D level set is applied to every slice of the segmented object to refine the 3D contours using the 3D level set contours as the initial contour. Further details on the level sets used can be found in the literature.[34] An example of the segmented bladder for CTU slice shown in Fig. 4 using the DL-CNN bladder likelihood map with level sets (DL-CNN with level sets) is shown in Fig. 9.

2.D. Bladder likelihood map generation using Haar features and random forest classifier

To compare the performance of DL-CNN for bladder likelihood map generation, the maps were also generated using Haar features and random forest classifier. Fifty-nine Haar features were extracted from the 32 × 32-pixels ROIs used to train the DL-CNN. A large number of Haar features can be extracted from a 32×32-pixels ROI. Using every possible Haar feature would be difficult due to the enormous number of features that would be generated; therefore, we considered the representative shapes for the bladder boundaries, and after experimenting on the training cases, we selected 59 different Haar features to generate the bladder likelihood maps, which are described in Table II.

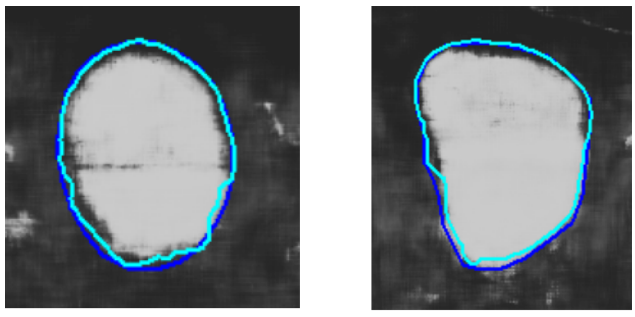
The extracted features were used to train a random forest classifier which combined the features together to generate a score that corresponds to an ROI's likelihood of being inside the bladder. The random forest classifier with 100 trees was trained using the same set of 160 000 training ROIs as described above for training the DL-CNN. The bladder likelihood map was generated by extracting the 59 Haar feature values from each ROI. The feature values were input into the trained random forest classifier, which output the likelihood that the input ROI was inside the bladder. The likelihood score for the ROI was assigned to the center pixel of the ROI. The collection of likelihood scores over the voxels in the VOI formed the bladder likelihood map.

The distribution of the Haar-feature-based bladder likelihood scores was different than that from the DL-CNN scores; thus a different threshold of 0.56 was chosen experimentally using the training cases and used to generate the binary bladder mask for initialization of the level sets. After the Haar-feature-based bladder binary mask was generated, the bladder segmentation process was identical to that described in Sec. 2.C.



(a) (b)

FIG. 10. Examples of bladder segmentations using DL-CNN with level sets for two cases in the test set. (a) Malignant bladder wall thickening was fully enclosed within the segmentation. (b) The bladder segmentation enclosed the lesion present in the bladder; however, the bottom of the contrast-enhanced region was slightly undersegmented. Arrows point to the wall thickening and lesion in (a) and (b), respectively. The lighter contour represents segmentation result from DL-CNN with level sets. The darker contour represents the radiologist's hand-outline.



(a) (b)

FIG. 11. Bladder likelihood maps and the corresponding bladder segmentation for cases shown in Fig. 10. (a) Refining the initial contour generated from the likelihood map by level sets results in accurate bladder segmentation. (b) Regions within the noncontrast region of the bladder had low likelihood of being within the bladder. The level sets propagated the initial contour to enclose the lesion and the noncontrast region. The lighter contour represents segmentation result from DL-CNN with level sets. The darker contour represents the radiologist's hand-outline.

TABLE III. Segmentation evaluation results using DL-CNN-based likelihood map with level sets averaged over the 81 training cases and 92 test cases.

	Volume intersection ratio (%) R^{3D}	Volume error (%) E^{3D}	Absolute volume error (mm) $\lvert E^{3D}\rvert$	Average minimum distance (%) AVDIST	Jaccard index (%) JACCARD^{3D}
Training set	87.2 ± 6.1	6.0 ± 9.1	8.8 ± 6.4	3.0 ± 1.2	81.9 ± 7.6
Test set	81.9 ± 12.1	10.2 ± 16.2	14.0 ± 13.0	3.6 ± 2.0	76.2 ± 11.8

2.E. Evaluation methods

Segmentation performance was evaluated by comparing the automatic segmentation results to the 3D hand-segmented contours. The volume intersection ratio, the volume error, the average minimum distance, and the Jaccard index[35] between the hand-segmented contours and computer-segmented contours were calculated. The performance metrics are described briefly below and more details can be found in our previous studies.[17,18]

The volume intersection ratio (R^{3D}) is the ratio of the intersection between the reference volume and the given volume to the reference volume,

$$R^{3D} = \frac{V_G \cap V_U}{V_G}, \tag{6}$$

where V_G is the volume enclosed by the reference standard contour G and V_U is the volume enclosed by the contour U being evaluated.

The volume error (E^{3D}) is the ratio of the difference between the reference volume and the given volume to the reference volume,

$$E^{3D} = \frac{V_G - V_U}{V_G}, \tag{7}$$

where positive error indicates undersegmentation, whereas a negative error indicates oversegmentation. Because the average of the volume error does not show the actual deviations from the reference standard due to oversegmentation and undersegmentation, the absolute error $|E^{3D}|$ is also calculated. From the volume intersection ratio and the volume error, other performance indicators can be derived.[36]

The average distance, AVDIST, is the average of the distances between the closest points of the two contours,

$$\mathrm{AVDIST}(G,U) = \frac{1}{2}\left(\frac{\sum_{x\in G}\min\{d(x,y): y\in U\}}{N_G} + \frac{\sum_{y\in U}\min\{d(x,y): x\in G\}}{N_U}\right), \tag{8}$$

where G and U are two contours being compared. N_G and N_U denote the number of voxels on G and U, respectively. The function d is the Euclidean distance. For a given voxel along the contour G, the minimum distance to a point along the contour U is determined. The minimum distances obtained for all points along G are averaged. This process is repeated by switching the roles of G and U. AVDIST is then calculated as the average of the two average minimum distances.

The Jaccard index (JACCARD^{3D}) is defined as the ratio of the intersection between the reference volume and the segmented volume to the union of the reference volume and the segmented volume,

$$\mathrm{JACCARD}^{3D} = \frac{V_G \cap V_U}{V_G \cup V_U}. \tag{9}$$

A value of 1 indicates that V_U completely overlaps with V_G, whereas a value of 0 implies that V_U and V_G are disjoint.

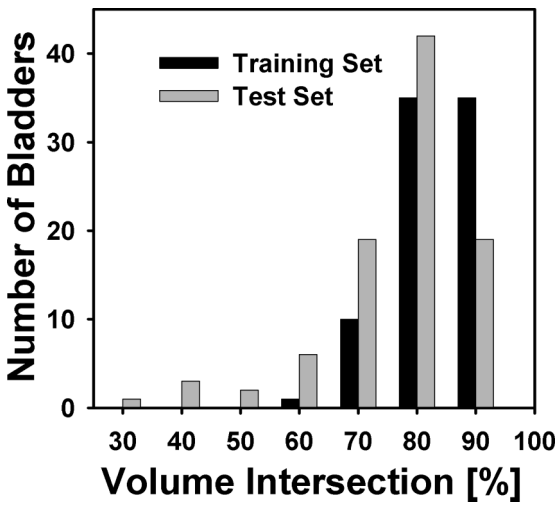


FIG. 12. Histogram of the percent volume intersection ratio for the training and test sets. The mean volume intersection was 87.2% for the 81 training cases and 81.9% for the 92 test cases.

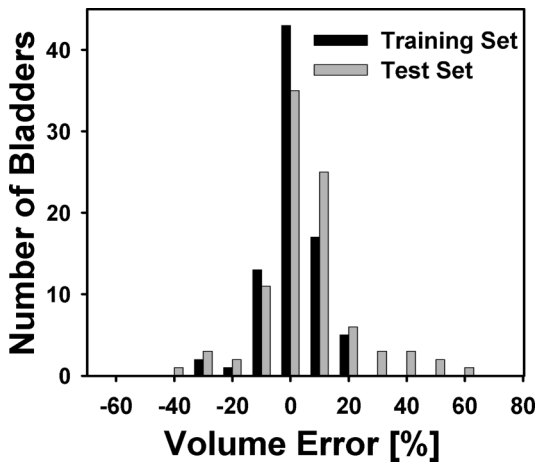


FIG. 13. Histogram of the volume error for the training and test sets. The mean volume error was 6.0% for the 81 training cases and 10.2% for the 92 test cases.

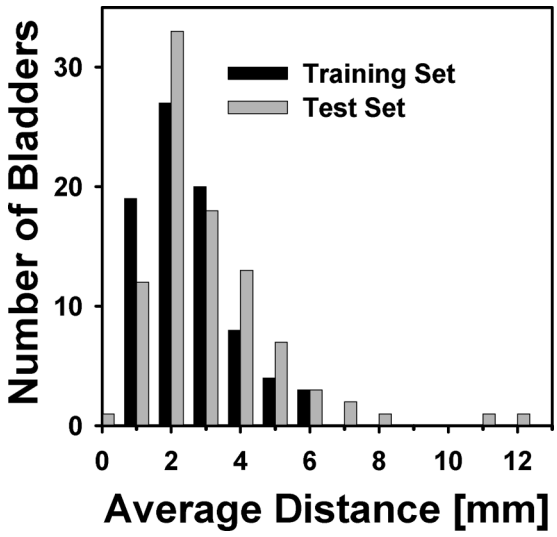


FIG. 14. Histogram of the average distance for the training and test sets. The mean average distance was 3.0 mm for the 81 training cases and 3.6 mm for the 92 test cases.

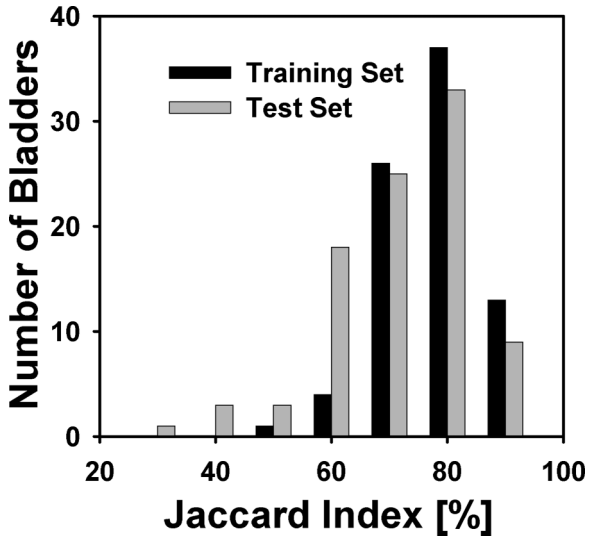


FIG. 15. Histogram of the Jaccard index for the training and test sets. The mean Jaccard index was 81.9% for the 81 training cases and 76.2% for the 92 test cases.

3. RESULTS

3.A. Segmentation performance using DL-CNN bladder likelihood map with level sets

The trained DL-CNN obtained a classification error rate of 0.054 for the training set. The error rate for the classification of the ROIs was not measured, as the classification of the ROI is not the final goal of this study. Examples of the segmentation from cases in the test set are shown in Fig. 10. Figure 11 shows the bladder likelihood maps used to generate the bladder boundaries in Fig. 10. The segmentation performance measures averaged over the cases in the training and test sets are presented in Table III.

The histograms for volume intersection ratio, volume error, average distance, and the Jaccard index for both the training set and the test set are shown in Figs. 12–15, respectively.

Of the 81 cases in the training set, 70 bladders (86.4%) had a volume intersection ratio greater than 80% (Fig. 12). There were 79 bladders (97.5%) whose absolute volume error for the training set was less than 20% (Fig. 13). Forty-six bladders (56.8%) in the training set had an average distance less than 3 mm (Fig. 14), and 50 bladders (61.7%) had Jaccard indices of over 80% (Fig. 15).

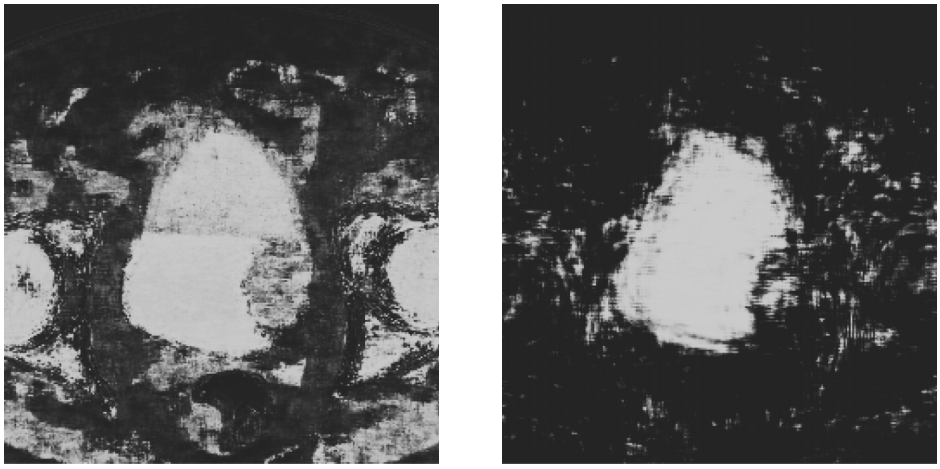
Of the 92 test cases, 61 bladders (66.3%) had a volume intersection ratio greater than 80% (Fig. 12). There were 73 bladders (79.3%) whose absolute volume error for the test set was less than 20% (Fig. 13). Forty-six bladders (50.0%) in the test set had an average distance less than 3 mm (Fig. 14) and 42 bladders (45.7%) had Jaccard indices of over 80% (Fig. 15).

3.B. Dependence of segmentation performance on input ROI size and DL-CNN pooling

Table IV summarizes the segmentation performance on the test cases for the conditions: (1) the maximum pooling layers were replaced by average pooling layers while keeping the input ROI size at 32×32 pixels and other parameters are the same as those in Sec. 3.A and (2) the input ROI size was changed to 16×16 and 64×64 pixels while all other parameters are the same as those in Sec. 3.A. The training set results showed similar trends. Figure 16 shows examples of the bladder likelihood map for 16×16-pixels ROI and the 64×64-pixels ROI for the CTU slice shown in Fig. 4.

TABLE IV. Segmentation evaluation results for DL-CNN with level sets using average pooling with 32×32-pixels ROI, and maximum pooling using 16×16-pixels ROI, and 64×64-pixels ROI averaged over the 92 test cases. Training set results showed similar trends.

	Volume intersection ratio (%) R^{3D}	Volume error (%) E^{3D}	Absolute volume error (%) $\lvert E^{3D} \rvert$	Average minimum distance (mm) AVDIST	Jaccard index (%) JACCARD^{3D}
Average pooling 32×32-pixels ROI	81.0 ± 12.1	5.3 ± 21.5	16.2 ± 14.9	4.5 ± 2.9	72.1 ± 13.3
Maximum pooling 16×16-pixels ROI	79.2 ± 14.2	11.0 ± 20.1	17.4 ± 14.8	4.4 ± 2.5	72.6 ± 14.0
Maximum pooling 64×64-pixels ROI	67.1 ± 12.7	24.9 ± 19.8	27.9 ± 15.1	6.4 ± 2.8	62.8 ± 13.1



(a) (b)

FIG. 16. Bladder likelihood map of the CTU slice shown in Fig. 4 using different ROI sizes. (a) Likelihood map generated using 16×16-pixels ROIs. (b) Likelihood map generated using 64 ×64-pixels ROIs.

3.C. Variability of reference standards

Table V shows the segmentation results using DL-CNN likelihood map with level sets compared against the two reference standards, as well as the results comparing the two hand-outlines with each other.

3.D. Comparison of segmentation performance using DL-CNN-based and Haar-feature-based bladder likelihood maps

Table VI summarizes the segmentation performance measures using the Haar-feature-based likelihood map to guide the level sets, averaged over the cases in the training and test sets. An example comparing the segmented bladder using the Haar-feature-based likelihood map with that using the DL-CNN-based likelihood map is shown in Fig. 17.

Table VII shows the initial segmentation surface ($\phi_0(x)$) generated from the DL-CNN-based and Haar feature-based bladder likelihood maps in comparison to the hand-outlines (RS1). The results show the segmentation performance without the refinement by the level sets and the differences between the DL-CNN-based likelihood maps and the Haar-feature-based likelihood maps.

3.E. Comparison of segmentation performance using DL-CNN bladder likelihood map with level sets and CLASS with LCR

Segmentation results of several test cases for both CLASS with LCR and DL-CNN with level sets are shown in Fig. 18. The segmentation performance measures for CLASS with LCR method are shown in Table VIII.

4. DISCUSSION

In this study, a new segmentation method that combines a likelihood map generated by DL-CNN with cascading level sets was developed and applied to a data set containing bladders in CTUs having a wide range of image quality. Most of the bladders were partially filled with excreted contrast material; however, some bladders were entirely filled with excreted contrast material and others did not contain any contrast-enhanced urine due to variation in timing for image acquisition. The presence of the two distinct areas that have very different attenuation values: an area filled with contrast material and an area without contrast material pose a challenge for segmentation that needs to go across the strong boundary. To alleviate this problem, we previously

TABLE V. Segmentation evaluation results in a subset of test cases with lesions (41 training cases, 50 test cases) between hand-segmented reference standards (RS1, RS2) by two different readers for DL-CNN with level sets. Segmentation evaluation of RS2 using RS1 as the reference is included to show interobserver variations.

		Volume intersection ratio (%) R^{3D}	Volume error (%) E^{3D}	Absolute volume error (%) $\lvert E^{3D}\rvert$	Average minimum distance (mm) AVDIST	Jaccard index (%) $JACCARD^{3D}$
DL-CNN vs RS1	Training set	85.9 ± 6.6	6.9 ± 9.6	9.3 ± 7.1	3.2 ± 1.3	80.4 ± 8.4
	Test set	81.2 ± 11.5	12.5 ± 13.5	13.4 ± 12.5	3.6 ± 1.9	76.4 ± 11.5
DL-CNN vs RS2	Training set	84.3 ± 7.1	9.7 ± 10.0	11.4 ± 7.9	3.4 ± 1.3	79.8 ± 8.2
	Test set	78.2 ± 10.9	17.5 ± 12.0	17.7 ± 11.6	4.0 ± 2.1	75.1 ± 11.0
RS2 vs RS1	Training set	96.2 ± 2.8	−3.0 ± 4.8	4.2 ± 3.8	1.4 ± 0.5	90.2 ± 4.8
	Test set	95.0 ± 8.1	−6.2 ± 15.3	10.3 ± 12.8	1.7 ± 1.0	86.1 ± 9.5

TABLE VI. Segmentation evaluation results using Haar-feature-based likelihood map with level sets averaged over 81 training cases and 92 test cases.

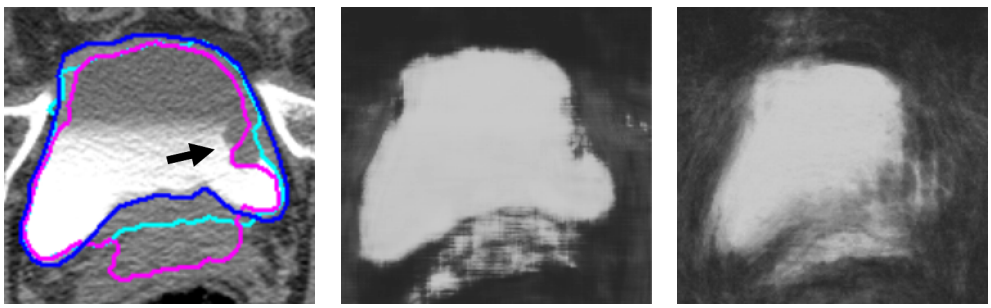
	Volume intersection ratio (%) R^{3D}	Volume error (%) E^{3D}	Absolute volume error (%) $\lvert E^{3D}\rvert$	Average minimum distance (mm) AVDIST	Jaccard index (%) $JACCARD^{3D}$
Training set	76.2 ± 10.4	15.5 ± 15.0	18.1 ± 11.6	5.2 ± 1.7	70.7 ± 10.0
Test set	74.3 ± 12.7	13.0 ± 22.3	20.5 ± 15.7	5.7 ± 2.6	66.7 ± 12.6

require two manually input VOIs: one for the noncontrast region and the other for the contrast-enhanced region using our CLASS segmentation method[17,18] and an LCR method was needed to refine and connect the two contours. However, by combining the DL-CNN bladder likelihood maps with the level set methods, we no longer needed the separate input user inputs for the two different regions. A major contribution of this work is that it demonstrates the DL-CNN can overcome the strong boundary between two regions that have large difference in gray levels and provides a seamless mask to guide level set segmentation. This has been a problem for many gradient-based segmentation methods. As a result, this new method requires only one user input bounding box for the entire bladder to start the segmentation procedure compared to the two user input bounding boxes for the previous method.

Compared to our CLASS with LCR method using the same data set, segmentation using DL-CNN performed better. All performance measures were improved using DL-CNN with level sets compared to CLASS with LCR for both the training and test sets. The differences in the volume intersection ratio, absolute volume error, average minimum distance, and the Jaccard index for the training set were statistically significant, with p-values of 0.01, 0.007, 0.01, and 0.002, respectively, by two-tailed paired t-test at an alpha level of 0.01 after the Bonferroni correction for the five comparisons. For the test set, the differences in the volume intersection ratio, volume error, and the absolute volume error were statistically significant with p-values of 0.004, 0.001, and 0.005 by two-tailed paired t-test at the alpha level of 0.01. For the training set, the percentages of cases obtaining improvements in the five performance improvements in the five performance measures range from 54% to 64%. For the test set, the improvements range from 54% to 67%. More importantly, DL-CNN with level sets better included lesions within its segmented region; 50 out of 59 (84.7%) lesions in the training set and 64 out of 78 (82.1%) lesions in the test set were included better than or similar to the bladder segmented with CLASS with LCR. These improvements were obtained while reducing the number of user inputs (one box vs two boxes).

DL-CNN with level sets generally enclosed more of the lesions within the segmented regions compared to CLASS with LCR [Figs. 18(a) and 18(b)], which is important because further steps of the CAD system for lesion detection and characterization will be performed within the segmented bladder. The noncontrast-enhanced region was segmented more accurately, without leaking into the adjacent organs using the DL-CNN [Figs. 18(a) and 18(c)]. However, there were cases that performed worse than our previous method [Fig. 18(d)]. These were caused by either the network giving low likelihood scores for portions of the bladder, causing undersegmentation, or the network giving relatively high likelihood scores for other organs, such as the bone, causing oversegmentation. Organs that were given relatively high bladder likelihood scores, such as the femoral heads, can be seen in the regions outside the VOI in Fig. 7.

For a few test set cases, DL-CNN with level sets performed well below the average performance of the test data set. Some of these cases had poor image quality due to noise caused by the large patient size or the presence of hip prosthesis. Other large segmentation mistakes were due to the patient having advanced bladder cancer spreading into the neighboring organs and causing the segmentation to leak into those areas.



(a) (b) (c)

FIG. 17. Comparison of bladder segmentations using DL-CNN-based likelihood map and Haar-feature-based likelihood map. (a) DL-CNN-based segmentation (light contour) encloses the bladder lesion within the segmentation, while the Haar-feature-based segmentation (darker contour) does not fully enclose the lesion and leaks into the prostate. The arrow points to the lesion. The darkest contour represents the radiologist's hand-outline. (b) Bladder likelihood map generated using DL-CNN. (c) Bladder likelihood map generated using Haar features and random forest classifier.

TABLE VII. Segmentation evaluation results using initial contours (no level sets) generated using bladder likelihood maps with DL-CNN and Haar features averaged over the 92 test cases. Training cases showed similar trends.

	Volume intersection ratio (%) R^{3D}	Volume error (%) E^{3D}	Absolute volume error (%) $\|E^{3D}\|$	Average minimum distance (mm) AVDIST	Jaccard index (%) $JACCARD^{3D}$
DL-CNN	68.7 ± 12.0	27.3 ± 13.7	27.4 ± 13.6	5.7 ± 2.2	66.2 ± 11.8
Haar features	59.8 ± 12.1	32.3 ± 18.6	34.0 ± 15.2	8.1 ± 2.6	55.6 ± 11.4

We are working on improving our method to reduce the errors caused by these types of cases.

When average pooling was used instead of maximum pooling in the network structure, the segmentation performance measures deteriorate in general. The differences between the two methods were statistically significant for the volume error, average minimum distance, and the Jaccard index for the test set.

Using the 16×16-pixels ROIs as input to the network resulted in bladder likelihood maps with finer details, such as lesion boundaries and the boundary between the noncontrast and the contrast-enhanced regions of the bladder. However, these maps did not lead to better segmentation than the likelihood maps obtained from 32×32-pixels ROIs likely due to the fine details hindering the generation of the initial contour for the entire bladder. On the other hand, the bladder likelihood maps obtained from 64×64-pixels ROIs contained less details from the structures surrounding the bladder. However, the shapes of the bladder might have lost too much details compared to those in the likelihood maps generated using 32×32-pixels ROIs. It also had the tendency of misclassifying large lesions as outside of the bladder. As shown in Table IV, both the smaller 16×16-pixels ROI and the larger 64×64-pixels ROIs were inferior to the 32×32-pixels ROI for generating the bladder likelihood maps to guide bladder segmentation.

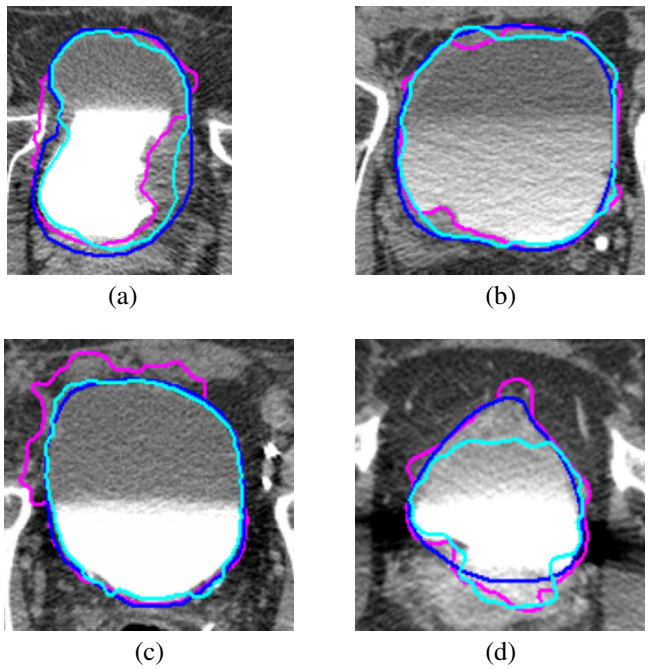


FIG. 18. Comparison of bladder segmentation using DL-CNN with level sets and CLASS with LCR. (a) DL-CNN slightly undersegments the upper region of the noncontrast region but encloses more of the large, malignant lesion and does not leak toward the bones. (b) The two segmentation methods perform similarly, but DL-CNN with level sets encloses the lesion, whereas CLASS does not. (c) DL-CNN with level sets does not leak into the surrounding organs in the noncontrast region, unlike CLASS. (d) CLASS performs better than DL-CNN with level sets in the noncontrast-enhanced region. Both methods oversegment the contrast-enhanced region. The light contour represents segmentation using DL-CNN with level sets. The darker contour represents segmentation using CLASS with LCR. The darkest contour represents the radiologist's hand-outline.

The bladder segmentation using DL-CNN with level sets performed comparably regardless of which of the two hand-outlines was used as the reference standard. The agreement between the computer and the hand-outlines is slightly lower than the agreement between the two observers (approximately 10% for the volume intersection ratio and the Jaccard index), but the computer segmentation in this range of accuracy is acceptable and still useful for defining the search region for bladder lesion detection, as shown in our previous work on bladder lesion detection.[9]

Comparing Tables III and VI, it is seen that the bladder likelihood maps obtained from the Haar features and the random forest classifier were not as effective as those from the DL-CNN, resulting in lower bladder segmentation performance. The differences in all performance measures but the volume error for the test set were statistically significant.

The comparison of the initial segmentation surfaces generated from the bladder likelihood maps with the reference standards shows that the DL-CNN-based maps are closer to the hand-outlines than the Haar-feature-based maps (Table VII). The result also shows that segmentation using the DL-CNN alone cannot reach the high performance level achieved by DL-CNN with refinement by level sets. The DL-CNN bladder likelihood maps are generally undersegmenting the bladder, often catching the edge of the inner bladder wall for cases with circumferential bladder wall thickening while lowering the threshold for the DL-CNN bladder likelihood map would lead to leaking. Applying the level sets to the slightly undersegmented contours allows better control of the balance between under and oversegmentation.

We chose the network structure size and level set parameters by experimentation where each parameter was varied over a reasonable range, and the best parameter within the studied range was chosen based on the evaluation of the training set results. Our sensitivity analysis of the level sets can be found

TABLE VIII. CLASS with LCR segmentation evaluation results averaged over the 81 training cases and 92 test cases.

	Volume intersection ratio (%) R^{3D}	Volume error (%) E^{3D}	Absolute volume error (%) $\|E^{3D}\|$	Average minimum distance (mm) AVDIST	Jaccard index (%) $JACCARD^{3D}$
Training set	84.2 ± 11.4	8.2 ± 17.4	13.0 ± 14.1	3.5 ± 1.9	78.8 ± 11.6
Test set	78.0 ± 14.7	16.5 ± 16.8	18.2 ± 15.0	3.8 ± 2.3	73.9 ± 13.5

in the literature.[34] We have performed a sensitivity analysis of the network structure size. The number of kernels within the first two convolution layers was varied between 32, 64, and 96. The network was trained on the training set, and the bladders were segmented using DL-CNN likelihood maps with level sets. The change in the volume intersection ratio was in the range of 0.5%–1.9%, absolute volume error 0.2%–9.3%, average minimum distance 0.6%–10.1%, and the Jaccard index 0.1%–2.2%. These results demonstrate that our DL-CNN based segmentation system is robust within a reasonable range of parameters.

A limitation to the new method is the long training time for the DL-CNN. The DL-CNN requires training, which takes approximately 5.5 h for 160 000 ROIs and 1500 iterations. However, the processes involving the DL-CNN have not been optimized, and a slower GPU was used for compatibility reasons for this study. Optimizing the process and using faster hardware will reduce the runtime for training the DL-CNN. Once the DL-CNN has been trained, it takes approximately 4 min to generate the bladder likelihood maps within the VOI of a case. It takes 2–5 min to mark the VOI and run the level set segmentation, depending on the bladder size. On the other hand, CLASS with LCR takes approximately 4–10 min per case to mark the VOI and run the segmentation. Therefore, for an unknown case, it may require up to 10 min for DL-CNN with level sets, which is comparable to the CLASS with LCR method.

It is difficult to perform direct comparison of segmentation performance to the previous methods by other investigators summarized in the Introduction due to the differences in the data sets and in their degrees of difficulty. A rough comparison can be made to only one of the studies,[14] which reported quantitative results. Chai *et al.*[14] achieved Jaccard indices of 70.5% and 77.7% for their automatic and semiautomatic methods, respectively, using 95 scans of 8 patients for training, and 233 scans of 22 patients for testing. Our segmentation method using DL-CNN achieved higher accuracy than the automatic method from Chai *et al.*[14] and achieved comparable results to their semiautomatic method, while using a larger independent test set.

5. CONCLUSION

Our results show that the proposed segmentation method using DL-CNN can accurately segment the bladders on CTU scans. While only using a single bounding box for the entire bladder as the input to the system, the new method performed comparable to or better than our previous CLASS with LCR method for all performance measures, which required two bounding boxes as input. However, the cost of this improvement is the increased runtime for training the DL-CNN. Once the DL-CNN is trained and implemented as a part of the segmentation package, the runtime for an unknown case becomes comparable. We observed that DL-CNN can differentiate the inside and outside of the bladder regions better than the Haar features with random forest classifier, resulting in a more accurate bladder likelihood map and segmentation after refinement by level sets. Further work is underway to optimize the segmentation process and to improve the segmentation accuracy, and especially important is to include the bladder lesions inside the segmented bladder boundaries. This study is a step toward the development of a reliable system for segmentation of bladders, which is a critical component of a CAD system for detection of urothelial lesions imaged with CT urography.

ACKNOWLEDGMENT

This work is supported by USPHS Grant No. U01CA179106.

[a]Author to whom correspondence should be addressed. Electronic mail: heekon@med.umich.edu; Telephone: (734) 647-8556; Fax: (734) 615-5513.

[1]American Cancer Society, What are the key statistics about bladder cancer?, 2015, available at www.cancer.org.
[2]S. A. Akbar, K. J. Mortele, K. Baeyens, M. Kekelidze, and S. G. Silverman, "Multidetector CT urography: Techniques, clinical applications, and pitfalls," Semin. Ultrasound CT MRI **25**, 41–54 (2004).
[3]E. M. Caoili, R. H. Cohan, M. Korobkin, J. F. Platt, I. R. Francis, G. J. Faerber, J. E. Montie, and J. H. Ellis, "Urinary tract abnormalities: Initial experience with multi-detector row CT urography," Radiology **222**, 353–360 (2002).
[4]W. C. Liu, K. J. Mortele, and S. G. Silverman, "Incidental extraurinary findings at MDCT urography in patients with hematuria: Prevalence and impact on imaging costs," Am. J. Roentgenol. **185**, 1051–1056 (2005).
[5]C. L. McCarthy and N. C. Cowan, "Multidetector CT urography (MD-CTU) for urothelial imaging," Radiology **225**, 237 (2002).
[6]M. Noroozian, R. H. Cohan, E. M. Caoili, N. C. Cowan, and J. H. Ellis, "Multislice CT urography: State of the art," Br. J. Radiol. **77**, S74–S86 (2004).
[7]S. B. Park, J. K. Kim, H. J. Lee, H. J. Choi, and K.-S. Cho, "Hematuria: Portal venous phase multi detector row CT of the bladder–a prospective study," Radiology **245**, 798–805 (2007).
[8]G. S. Sudakoff, D. P. Dunn, M. L. Guralnick, R. S. Hellman, D. Eastwood, and W. A. See, "Multidetector computerized tomography urography as the primary imaging modality for detecting urinary tract neoplasms in patients with asymptomatic hematuria," J. Urol. **179**, 862–867 (2008).
[9]K. Cha, L. Hadjiiski, H.-P. Chan, R. H. Cohan, E. M. Caoili, and C. Zhou, "Detection of urinary bladder mass in CT urography with SPAN," Med. Phys. **42**, 4271–4284 (2015).

[10]L. Li, Z. Wang, X. Li, X. Wei, H. L. Adler, W. Huang, S. Rizvi, H. Meng, D. P. Harrington, and Z. Liang, "A new partial volume segmentation approach to extract bladder wall for computer aided detection in virtual cystoscopy," Proc. SPIE **5369**, 199–206 (2004).
[11]C. Duan, Z. Liang, S. Bao, H. Zhu, S. Wang, G. Zhang, J. J. Chen, and H. Lu, "A coupled level set framework for bladder wall segmentation with application to MR cystography," IEEE Trans. Med. Imaging **29**, 903–915 (2010).
[12]C. J. Duan, K. H. Yuan, F. H. Liu, P. Xiao, G. Q. Lv, and Z. R. Liang, "An adaptive window-setting scheme for segmentation of bladder tumor surface via MR cystography," IEEE Trans. Inf. Technol. Biomed. **16**, 720–729 (2012).
[13]H. Han, L. Li, C. Duan, H. Zhang, Y. Zhao, and Z. Liang, "A unified EM approach to bladder wall segmentation with coupled level-set constraints," Med. Image Anal. **17**, 1192–1205 (2013).
[14]X. F. Chai, M. van Herk, A. Betgen, M. Hulshof, and A. Bel, "Automatic bladder segmentation on CBCT for multiple plan ART of bladder cancer using a patient-specific bladder model," Phys. Med. Biol. **57**, 3945–3962 (2012).
[15]L. Hadjiiski, H. P. Chan, Y. Law, R. H. Cohan, E. M. Caoili, H. C. Cho, C. Zhou, and J. Wei, "Segmentation of urinary bladder in CT urography (CTU) using class," Proc. SPIE **8315**, 83150J-83151–83150J-83157 (2012).
[16]L. M. Hadjiiski, B. Sahiner, H. P. Chan, E. M. Caoili, R. H. Cohan, and C. Zhou, "Automated segmentation of urinary bladder and detection of bladder lesions in multi-detector row CT urography," Proc. SPIE **7260**, 72603R-72601–72603R-72607 (2009).
[17]L. Hadjiiski, H. P. Chan, R. H. Cohan, E. M. Caoili, Y. Law, K. Cha, C. Zhou, and J. Wei, "Urinary bladder segmentation in CT urography (CTU) using CLASS," Med. Phys. **40**, 111906 (10pp.) (2013).
[18]K. Cha, L. M. Hadjiiski, H.-P. Chan, E. M. Caoili, R. H. Cohan, and C. Zhou, "CT urography: Segmentation of urinary bladder using CLASS with local contour refinement," Phys. Med. Biol. **59**, 2767–2785 (2014).
[19]H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," Med. Phys. **22**, 1555–1567 (1995).
[20]R. K. Samala, H. P. Chan, Y. Lu, L. M. Hadjiiski, J. Wei, and M. A. Helvie, "Digital breast tomosynthesis: Computer-aided detection of clustered microcalcifications on planar projection images," Phys. Med. Biol. **59**, 7457–7477 (2014).
[21]M. N. Gurcan, B. Sahiner, H. P. Chan, L. M. Hadjiiski, and N. Petrick, "Selection of an optimal neural network architecture for computer-aided diagnosis—Comparison of automated optimization techniques," Radiology **217**, 436 (2000).
[22]M. N. Gurcan, B. Sahiner, H. P. Chan, L. M. Hadjiiski, and N. Petrick, "Selection of an optimal neural network architecture for computer-aided detection of microcalcifications—Comparison of automated optimization techniques," Med. Phys. **28**, 1937–1948 (2001).
[23]J. Ge, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, J. Wei, M. A. Helvie, and C. Zhou, "Computer aided detection of clusters of microcalcifications on full field digital mammograms," Med. Phys. **33**, 2975–2988 (2006).
[24]J. Ge, L. M. Hadjiiski, B. Sahiner, J. Wei, M. A. Helvie, C. Zhou, and H.-P. Chan, "Computer-aided detection system for clustered microcalcifications: Comparison of performance on full-field digital mammograms and digitized screen-film mammograms," Phys. Med. Biol. **52**, 981–1000 (2007).
[25]P. Filev, L. Hadjiiski, H.-P. Chan, B. Sahiner, J. Ge, M. A. Helvie, M. Roubidoux, and C. A. Zhou, "Automated regional registration and characterization of corresponding microcalcification clusters on temporal pairs of mammograms for interval change analysis," Med. Phys. **35**, 5340–5350 (2008).
[26]R. K. Samala, H. P. Chan, Y. Lu, L. M. Hadjiiski, J. Wei, and M. A. Helvie, "Computer- aided detection system for clustered microcalcifications in digital breast tomosynthesis using joint information from volumetric and planar projection images," Phys. Med. Biol. **60**, 8457–8479 (2015).
[27]A. Krizhevsky, CUDA-CONVNET, 2012, see https://code.google.com/p/cuda-convnet/.
[28]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *NIPS 2012: Neural Information Processing Systems*, *Lake Tahoe, Nevada* (2012).
[29]O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein, "Imagenet large scale visual recognition challenge," Int. J. Comput. Vis. **115**, 211–252 (2015).
[30]A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," M.S. thesis, University of Toronto, Toronto, 2009, see http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.
[31]P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, edited by A. Jacobs and T. Baldwin (IEEE Computer Society, Los Alamitos, CA, 2001), Vol. 1, pp. 511–518.
[32]R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proceedings of IEEE International Conference on Image Processing* (IEEE, 2002), Vol. I, pp. 900–903.
[33]V. Nair and G. E. Hinton, *Presented at the Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (2010).
[34]E. Street, L. Hadjiiski, B. Sahiner, S. Gujar, M. Ibrahim, S. K. Mukherji, and H. P. Chan, "Automated volume analysis of head and neck lesions on CT scans using 3D level set segmentation," Med. Phys. **34**, 4399–4408 (2007).
[35]P. Jaccard, "The distribution of the flora in the alpine zone," New Phytol. **11**, 37–50 (1912).
[36]T. W. Way, L. M. Hadjiiski, B. Sahiner, H.-P. Chan, P. N. Cascade, E. A. Kazerooni, N. Bogot, and C. Zhou, "Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours," Med. Phys. **33**, 2323–2337 (2006).