Automatic Detection and Segmentation of Lymph Nodes from CT Data

Adrian Barbu, Michael Suehling, Xun Xu, David Liu, S. Kevin Zhou, Dorin Comaniciu

Abstract—Lymph nodes are assessed routinely in clinical practice and their size is followed throughout radiation or chemotherapy to monitor the effectiveness of cancer treatment. This paper presents a robust learning-based method for automatic detection and segmentation of solid lymph nodes from CT data, with the following contributions. First, it presents a learning based approach to solid lymph node detection that relies on Marginal Space Learning to achieve great speedup with virtually no loss in accuracy. Second, it presents a computationally efficient segmentation method for solid lymph nodes (LN). Third, it introduces two new sets of features that are effective for LN detection, one that self-aligns to high gradients and another set obtained from the segmentation result. The method is evaluated for axillary LN detection on 131 volumes containing 371 LN, yielding a 83.0% detection rate with 1.0 false positive per volume. It is further evaluated for pelvic and abdominal LN detection on 54 volumes containing 569 LN, yielding a 80.0% detection rate with 3.2 false positives per volume. The running time is 5-20 seconds per volume for axillary areas and 15-40 seconds for pelvic. An added benefit of the method is the capability to detect and segment conglomerated lymph nodes.

Index Terms—lymph node detection, lymph node segmentation, cancer staging.

I. INTRODUCTION

Lymph node (LN) analysis is a difficult task that accounts for a significant part of daily clinical work in Radiology. In particular, automatic lymph node detection and segmentation is important for cancer staging and treatment monitoring. Lymph nodes nearby primary cancer regions are routinely assessed by clinicians to monitor disease progress and effectiveness of the cancer treatment. The assessment is usually based on 3D Computed Tomography (CT) data. When the cancer treatment is successful, the lymph nodes decrease in size. Since finding the lymph nodes is time consuming and highly dependent on the observer's experience, a system for automatic lymph node detection and measurement is desired. For follow-up studies, the system could further report the size change for each major lymph node.

In this paper we introduce a learning based method for the automatic detection and segmentation of solid lymph nodes. Enlarged lymph nodes with a solid interior are of particular clinical interest since they are believed to have a higher probability of being malignant than lymph nodes that for example have a fatty core.

A. Barbu is with the Department of Statistics, Florida State University, Tallahassee, FL 32306 USA. E-mail: abarbu@fsu.edu

M. Suehling, Xun Xu, D. Liu, S.K. Zhou and D. Comaniciu are with the Image Analytics and Informatics Department, Siemens Corporate Research, Princeton, NJ 08540, USA

According to the well-established RECIST guidelines [15] that are widely applied in clinical routine, lesions are considered to be measurable if they have a diameter of at least 10mm in CT. Only those measurable lesions will be considered as so-called target lesions to be recorded and tracked over time across follow-up exams during therapies. Following these guidelines, our method is targeted for detecting clinically relevant (and possibly malignant) lymph nodes of size at least 10mm. Similar requirements are confirmed in [4] for neck lymph nodes.
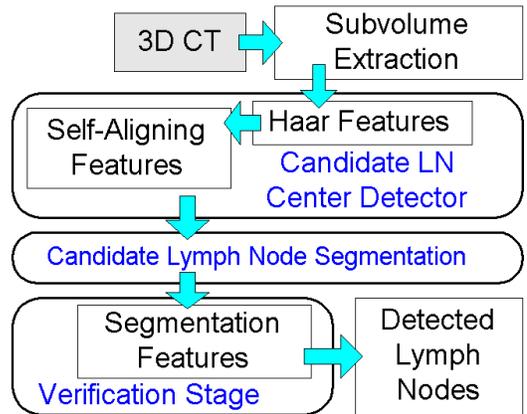

Fig. 1. Diagram of the proposed solid lymph node detection method.

The diagram of the proposed solid lymph node detection and segmentation method is shown in Fig. 1. For speed and accuracy, regions of interest (two axillary and one pelvic in our experiments) are extracted automatically as described in Section III-A. A number of lymph node center candidates are generated for each region using a two-stage detector described in Section III-B. Each candidate is segmented as described in Section III-C. Finally, a verification stage described in Section III-D gives the final result.

Earlier work appeared in [1], focused only on axillary LN detection. This work follows the same basic algorithm, and presents a more thorough evaluation of the LN detection on a larger axillary dataset as well as on another dataset focused on the pelvic/abdominal area. Moreover, we present an evaluation of the segmentation accuracy of the proposed approach.

II. RELATED WORK

There is a limited amount of work directed to automatic lymph node detection [5], [6], [7], [10], [21]. These works target mediastinal [6], [7], abdominal [10], pelvic [21] and neck [5] lymph nodes. Our work targets axillary and pelvic+abdominal lymph nodes.

While the work in [21] uses multiple MR sequences and special contrast, our work addresses lymph node detection in CT images. The proposed approach could in principle be adapted to other modalities such as MR or 3D ultrasound images, but that is subject to future work.

A special filter was used in [10] to detect lymph node centers. This minimum directional difference filter is constructed with the assumption that lymph nodes have uniform intensity and are spherical. The approach obtained a 57% detection rate with about 58 false positives per volume.

The Min-DD filter work [10] was improved in [6] by adding a Hessian-based blobness measure for reducing false positives. In contrast, our work comes from a learning perspective, where the appropriate features that discriminate the lymph nodes from other structures are learned from training examples. Blobness measures could also be included in the feature set of our system and will be selected by the learning algorithm if they are better than the other features. By using the learning based approach and highly discriminative features based on the LN segmentation, we observe a 30 times reduction in false positives for the same detection rate, when compared to [6].

Feulner et al [7] detect mediastinal lymph nodes using a learning-based lymph node detector based on Haar and steerable features [22]. Using location information based on automatic organ segmentation they obtain 64.8% detection rate with 2.9fp/vol. Our approach also uses Haar features to remove easy negatives, but it uses two novel sets of informative features, one aligned to high gradients and one based on LN segmentation to obtain state of the art results.

The idea of coupling segmentation with detection has been recently proposed in the computer vision literature [11]. Our work also combines segmentation with object detection, with the following differences.

1) Our segmentation method produces a defined object boundary whereas [11] has a fuzzy boundary.
2) Our work is oriented towards detecting 3D lymph nodes, which have a high degree of shape variability. In contrast, [11] detects 2D objects of specific shapes such as cars, cows and humans.
3) The segmentation is constructed differently in our work, using a Gaussian MRF and gradient descent as opposed to [11] where it is constructed in a greedy way from a number of patches.
4) Our work constructs segmentation-based features that are used to train a classifier, whereas [11] obtains a probability from the segmentation hypotheses by voting.

Semi-automatic lymph node segmentation in CT has been investigated in a number of papers [4], [20], [17], [19], using manual initialization of the lymph node location. In comparison, the approach proposed in this paper is fully automatic, the lymph nodes are detected and segmented without any user interaction. The proposed segmentation method can also be used with user initialization if desired.

A method based on a variant of level sets called fast marching was proposed in [20] for the 2D segmentation of LN on individual CT slices. No quantitative evaluation was reported.

In [4], a model based on springs that are arranged on a sphere triangulation is deformed outwards starting from a small initialization to find high gradient locations. The method was evaluated on 5 volumes containing 40 lymph nodes and obtained a smaller error (0.47mm) than our method (0.83-1mm) but requires manual initialization and is thousands of times slower than our segmentation method.

A novel graph based approach [19] uses a radial representation similar to ours, but with a different model with a cost function that encourages homogeneity and an algorithm based on the optimal surface reconstruction method of [12]. Results are similar to [4] hence better than our method, but require manual initialization.

Semi-automatic lymph node segmentation in multiple MRI images was presented in [17] where an ellipse model was evolved using PDEs.

III. Proposed Method for Solid Lymph Node Detection and Segmentation

The proposed lymph node detection and segmentation system first detects candidate lymph node centers using a learning based approach. Each candidate is used by a segmentation module to extract a candidate lymph node boundary. A learning-based verification stage uses features obtained from the data and the extracted boundary to score the candidates and keep only the best ones. The whole approach is summarized in Algorithm 1 below.

Algorithm 1 LN Detection and Segmentation

Input: CT volume at 1.5mm isotropic voxel size.
Output: Set D of detected and segmented lymph nodes.

1: Extract subvolume(s) of interest from the CT data (Section III-A).
2: **for** each subvolume V **do**
3:   Obtain initial set of candidates C_0 as all locations with intensity in $[-100, 200]$HU (Section III-B.1).
4:   Of the C_0 candidates, keep the candidates C_1 that pass the Haar detector (Section III-B.2).
5:   Of the C_1 candidates, keep the candidates C_2 that pass the self-aligning detector (Section III-B.3).
6:   **for** each $c_i = (x_i, y_i, z_i) \in C_2$ **do**
7:     Obtain a segmentation S_i with center c_i (Section III-C).
8:     Obtain a score p_i from the detector based on segmentation features extracted using S_i(Section III-D).
9:   **end for**
10:   Discard from C_2 all c_i with $p_i < \tau$, obtaining C_3.
11:   Call non-maximal suppression (Algorithm 2) on C_3 to obtain the detected lymph nodes D.
12: **end for**

The steps of the algorithm are explained in more detail in the rest of the section.

A. Subvolume of Interest Extraction

To constrain the search and for increased speed and accuracy, the two axillary regions and the pelvic region are

detected and cropped automatically. The axillary lymph node detection is performed on the two cropped axillary subvolumes and the pelvic lymph nodes in the pelvic subvolume. The axillary subvolumes are obtained by first detecting the lung tips, with an approach similar to [14]. This can be done very reliably and is not the object of this paper. Relative to the two lung tip locations (x, y, z), the axillary subvolumes have size $240 \times 240 \times 240$ mm and the upper-left corner at $(x+20, y-135, z-156)$ mm for the left lung and upper-right corner at $(x-20, y-135, z-156)$ mm for the right lung. The pelvic region is obtained by detecting the pubic symphysis tip and cropping a subvolume of size $280 \times 230 \times 200$ mm with the upper left corner at $(-120, -80, -90)$ mm relative to the detected pubic symphysis tip location.

At this stage we use fixed sizes for the ROIs. In the future the size of the ROIs can be adapted patient-specifically by using additional landmarks to define the bounding boxes.

B. Candidate Lymph Node Detection

Lymph node center candidates are detected in the cropped subvolumes ignoring the lymph node size and shape, in the spirit of Marginal Space Learning [22]. Marginal Space Learning is a method that speeds-up object detection by many orders of magnitude by first detecting object candidates in a subspace where some object parameters (e.g. size, orientation, etc) are ignored. This set of initial candidates is then refined by searching for some of the missing parameters using appropriate detectors.

In this spirit, the proposed approach first detects a set of LN center positions (x, y, z), thus ignoring the LN size and shape. The LN sizes and shapes are then obtained using the LN segmentation approach presented in Section III-C.

1. Initial set of candidates. The LN center positions are detected using a cascade of trained detectors starting from an initial set of LN positions containing all voxels in the cropped subvolumes that have an intensity in the interval [-100,200]HU. The candidates from this initial set are then evaluated using a fast detector based on Haar features followed by a second detector based on a newly proposed set of self-aligning gradient features. These two detectors are described in more detail below.

2. Haar Detector. The first detector is a cascade of Adaboost classifiers trained using 92,000 3D Haar features that have been described in [16]. For each candidate location, the Haar features are extracted from a window of size $23 \times 23 \times 23$ voxels centered at that location. This window size guarantees that most lymph nodes (up to 35mm diameter) will be inside the window.

Fig. 3. Self aligning features are computed along 14 directions relative to candidate position.

A cascade of detectors is trained using these features, with the parameters for Classifier 1 from Table I.

3. Self-Aligning Detector. The second detector uses a set of features that are self-aligned to high gradients. The self-alignment insures that the feature values will be consistent for different lymph nodes independent of their size and shape.
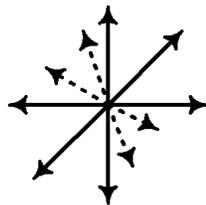
These features are computed based on rays casted in 14 directions in 3D space from each candidate location. These 14 directions are $(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)$, and $(\pm 1, \pm 1, \pm 1)$. Of the 14 directions, 10 are shown in Fig. 3 for clarity.

In each direction $d_i, 1 \leq i \leq 14$, local maxima of the gradient above each of 10 thresholds $\tau_j = 10j, 1 \leq j \leq 10$ (see Figure 4), are found at three scales $s_k = 1/2^k, 1 \leq k \leq 3$.

Some of the features are based on the 24 point features that were described in [22]. These 24 features are computed at a point (x, y, z) based on a given direction $d = (d_x, d_y, d_z)$ and are:

Fig. 4. In each direction, local gradient maxima above different thresholds τ_j are found.

- Intensity features $I^k(x, y, z), k \in \{1/3, 1/2, 1, 2, 3\}$ and $\ln I(x, y, z)$ where $I(x, y, z)$ is the intensity value
- The three components of the intensity gradient $g = (g_x(x, y, z), g_y(x, y, z), g_z(x, y, z))$
- The norm $\|g\|$ of the intensity gradient
- The dot product $g \cdot d$, $\ln |g \cdot d|$ and $|g \cdot d|^k$, $k \in \{1/3, 1/2, 1, 2, 3\}$
- The quantity $\sqrt{|\|g\|^2 - (g \cdot d)^2|}$
- The angle $\theta = \cos^{-1} \frac{g \cdot d}{\|g\|}$ as well as $|\theta|^k, k \in \{1/3, 1/2, 1, 2, 3\}$ and $\ln |\theta|$

The proposed self-aligning gradient features are the following:

- Each of the 24 point features described above is computed at each of the first three local maxima for each direction d_i, threshold τ_j and scale s_k.
- Each of the 24 features types described above is computed half way between the candidate location and each of the first three local maxima, for each d_i, τ_j, s_k.
- The distance to each of the first three local maxima for each d_i, τ_j, s_k.
- The differences between distances to the corresponding first three local maxima in any combination of two different directions d_i, d_j for each τ_k, s_l.

About 64,000 features are obtained from the 14 directions, three local maxima in each direction, three scales and the feature types described above. Some of these features are based on appearance while others are based on the shape of the iso-gradients around the LN candidate.

A cascade of detectors is trained using these features, with the parameters for Classifier 2 from Table I.

The best N_{max} candidates above a threshold are kept for each cropped subvolume. Examples of detected candidates are shown in Figure 2. For each candidate location, a segmentation is obtained as described below.

C. Candidate Lymph Node Segmentation

The segmentation algorithm is specially designed for detecting clinically highly relevant solid lymph nodes. The solid lymph nodes have a blob-like shape that can be described by using a radial function $r : S^2 \to \mathbb{R}$ defined on the sphere in 3D, representing the distance from the lymph node center to
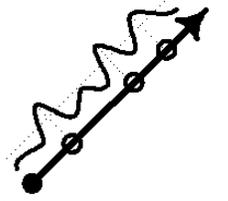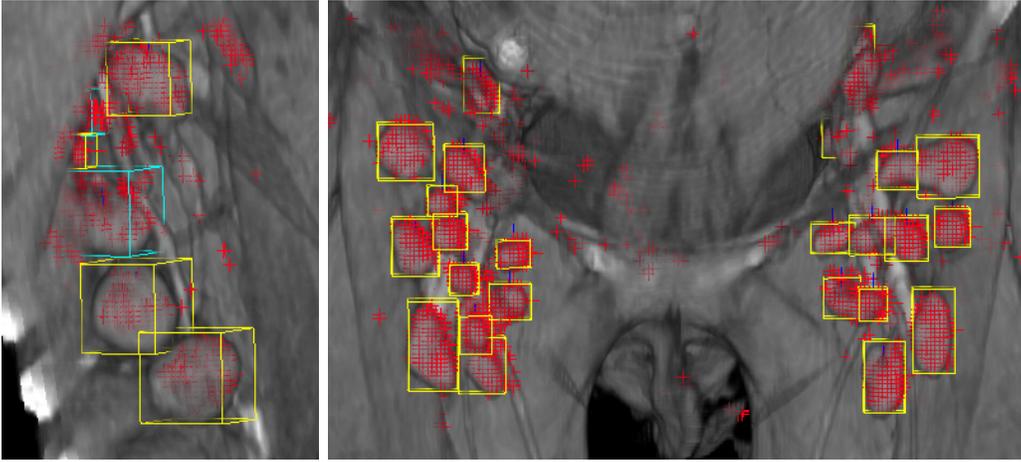
Fig. 2. Detected LN candidates shown as red crosses for an axillary region (left) and a pelvic region (right).

the boundary in all directions. The same shape representation was used in [19] for LN segmentation, but using a different model.

In this work, the sphere has been discretized using a triangulation with 162 vertices, 480 edges and 320 triangles, as shown in Figure 5. As the number of triangles of our sphere mesh can only be $5*4^k$, we chose 320 triangles to balance speed and accuracy for handling large lymph nodes. Alternatively, triangulations with 1280 or more triangles could also be used at a larger computational cost.
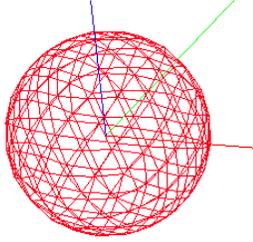

Fig. 5. Sphere triangulation with 162 vertices and 320 triangles.

The edges of this triangulation induce a neighborhood structure between the vertices. Two vertices are considered neighbors if there is a triangulation edge connecting them.

This shape representation can accurately describe blob-like shapes even when they are not convex. It has some difficulty representing the extremities of very elongated shapes (with aspect ratio at least 4:1), as shown in Figure 15. However, out of the more than 900 solid lymph nodes that have been manually segmented using this representation, we encountered such difficulties with 26 LN (2.8% of all solid LN).

Other examples of lymph node segmentations with this representation are shown in Figure 9, 12, 13 and 14. This representation is similar to the shape representation in [4].

Each of the 162 sphere vertices represents a direction $\mathbf{d}_i$. Given a candidate lymph node location $\mathbf{c}$ obtained by the candidate detector described in Section III-B, a segmentation using this location as the center is uniquely determined by the radii $r_i, i = 1, ..., N$ for all directions $\mathbf{d}_i$, where $N = 162$ in our case. These radii form a vector $\mathbf{r} = (r_1, ..., r_N)$.
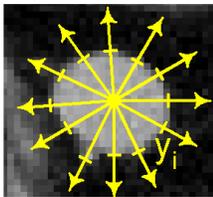

Fig. 6. Measurements y_i are found for each direction $\mathbf{d}_i$ as the most probable boundary location.

The lymph nodes exhibit high shape variability with no particular commonalities, making it inappropriate to describe them using generative models such as PCA. Instead of a PCA

model, we adopt a Gaussian MRF shape prior to constrain the shape of the obtained segmentation.

To find the segmentation vector $\mathbf{r}$ we propose an approach similar to the Active Shape Models [3], but using a robust data cost, gradient optimization and a Gaussian MRF shape prior.

Given the candidate location $\mathbf{c}$, the most likely lymph node boundary location y_i is found in each direction $\mathbf{d}_i, i = 1, ..., N = 162$ as the first location where the intensity difference from the candidate center is larger than a threshold D_{max}, as illustrated in Figure 6:

$$y_i = \arg \min_{r \in (0, R_{max})} |I(\mathbf{c}) - I(\mathbf{c} + r\mathbf{d}_i)| > D_{max}. \quad (1)$$

From the measurement vector $\mathbf{y} = (y_1, \ldots, y_N)$, the segmentation $\mathbf{r}$ is obtained by minimizing the following energy function

$$E(\mathbf{r}) = \alpha \sum_{i=1}^{N} \rho(r_i - y_i) + \sum_{i=1}^{N} \frac{1}{2|\partial i|} \sum_{j \in \partial i} (r_i - r_j)^2 \quad (2)$$

where $\rho(x) = \ln(1 + x^2/2)$ is a robust function and for each i, ∂i are the neighbors of i on the sphere triangulation. Remember that two vertices are neighbors if there is an edge of the sphere triangulation connecting them.

The first term in eq. (2) is a robust data term, while the second term is the Gaussian MRF prior that encourages the neighboring vertices to have similar radii. The robust data term ensures that the segmentation is robust to any sporadic outliers in the measurements y_i, fact illustrated in Figure 7.

Moreover, if a measurement y_i does not exist, its corresponding term is removed from the first sum of eq. (2).

Fig. 7. The robust data term and Gaussian MRF obtain a smooth segmentation that ignores outliers in the measurements y_i.

Minimization of the energy from eq. (2) is done by gradient descent, starting with $\mathbf{r} = \mathbf{y}$ as initialization. The energy gradient can be computed analytically, obtaining the following update iteration:

$$r_i \leftarrow r_i - \eta \left(\alpha \frac{r_i - y_i}{1 + (r_i - y_i)^2/2} + r_i - \frac{\sum_{j \in \partial i} r_j}{|\partial i|} \right) \quad (3)$$
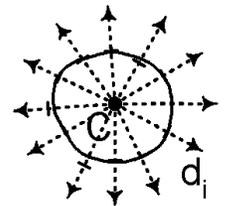
In practice, we use $\eta = 0.1$ and 1000 gradient update iterations, while $D_{max} = 50, \alpha = 1.6$. These parameters were chosen by cross-validation. The dependence of the detection performance on D_{max} and α is quite smooth, similar to the segmentation performance from Figure 18.

Other segmentation methods such as [4], [9], [20] could possibly be used, but they lack a robust data term, making them more prone to oversegmentations in low gradient locations.

D. Final Lymph Node Verification

For each of the candidate lymph node centers obtained using the candidate LN detector from Section III-B, a segmentation with 162 vertices is obtained as described in Section III-C. The segmentation is used to obtain more discriminative features for the final verification of the lymph node candidates.

For each lymph node, a bounding box is extracted from the segmentation and used to measure the LN size. Candidates whose second largest bounding box size is less than 9mm are automatically rejected purely based on their size since we are interested only in detecting lymph nodes larger than 10mm. We chose 9mm instead of 10mm as the rejection threshold in order to ac-

Fig. 8. Mesh vertices A and B are opposite if the line connecting them passes through the LN center C.

count for small errors in the automatic segmentation. The threshold could be changed accordingly to detect smaller lymph nodes if desired.

The LN verification is done using a trained detector based on the following features that are computed from the segmentation result:

- Each of the 24 point features described in Section III-B are computed at the 162 segmentation vertices using the directions from the LN center. For each feature, the 162 values are sorted in decreasing order.
- For each of the 24 point features, the 81 sums of feature values at the pairs of opposite vertices are computed and sorted in decreasing order. Two vertices are opposite if the line connecting them passes through the LN center, as illustrated in Figure 8. The sphere mesh is constructed to be symmetrical relative to the center, thus every vertex has exactly one opposite vertex.
- The 81 diameters (distances between opposite vertices relative to the segmentation center) are sorted in decreasing order. For each diameter the following features are computed:
  1) The size of each diameter.
  2) Asymmetry of each diameter, i.e. the ratio of the larger radius over the smaller radius.
  3) The ratio of the i-th sorted diameter and the j-th diameter for all $1 \le i < j \le 81$.
  4) For each of the 24 feature types, the max or min of the feature values at the two diameter ends.
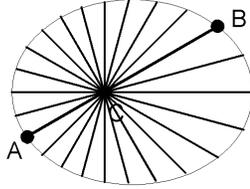 5) For each of the 24 feature types, the max or min of the feature values half way to the diameter ends.

In total there are about 17,000 features based on the shape and appearance of the segmentation result.

The verification detector is trained using the segmentation features described above, with the parameters for Classifier 3 from Table I. In this paper we present experiments using Adaboost [8], [13] and Random Forest [2], but other learning algorithms, e.g. Support Vector Machines [18] could also be used.

The trained classifier assigns a score p_i to each candidate i, a higher score representing a higher estimated likelihood to be a lymph node.

E. Non-maximal Suppression

All candidates with the score below a threshold τ are automatically removed. On the remaining candidates, a non-maximum suppression scheme is implemented as described in Algorithm 2.

The algorithm repeats adding the remaining candidate of highest score and removing all candidates close to it until no candidates are left.

Algorithm 2 Non-maximal Suppression

Input: Candidates $c_i = (x_i, y_i, z_i)$ with scores $p_i > \tau$ and bounding boxes b_i.
Output: Set D of detected lymph nodes.
1: Find the candidate c_i with highest score p_i.
2: **if** c_i exists **then** initialize $D = \{i\}$ **else** $D = \emptyset$, **stop**.
3: **while** true **do**
4:     Remove candidates c_j inside any box $b_i, i \in D$.
5:     Find remaining candidate c_j of highest score p.
6:     **if** c_j exists **then** add j to detected set: $D \leftarrow D \cup \{j\}$ **else stop**.
7: **end while**

Examples of detected and segmented lymph nodes are shown in red in Figure 9. The method can detect parts of lymph node conglomerates as shown in Figure 9, middle.

F. Training details

Table I contains the number of weak classifiers for each of the three detectors (based on Haar, Gradient and Segmentation features respectively). As the first two detectors contain a cascade of Adaboost classifiers, the number of weak classifiers of each level of the cascade are given in the corresponding table entries for both the axillary LN detector and the pelvic+abdominal LN detector. All parameters were chosen by cross-validation.

For comparison, a Random Forest [2] classifier with 100 trees was also trained on the segmentation-based features for the verification step.

IV. EXPERIMENTAL VALIDATION

The experimental results below are based on a six-fold cross-validation as follows. The CT volumes were divided into six disjoint sets. For each fold, the union of five of the sets was used to train the three classifiers and the remaining set was used for evaluation. Training all three classifiers for each fold takes about five hours.

TABLE I
TRAINING DETAILS FOR THE THREE CLASSIFIERS INCLUDING THE NUMBER OF WEAK CLASSIFIERS, DETECTION RATE AND FALSE POSITIVE RATE.

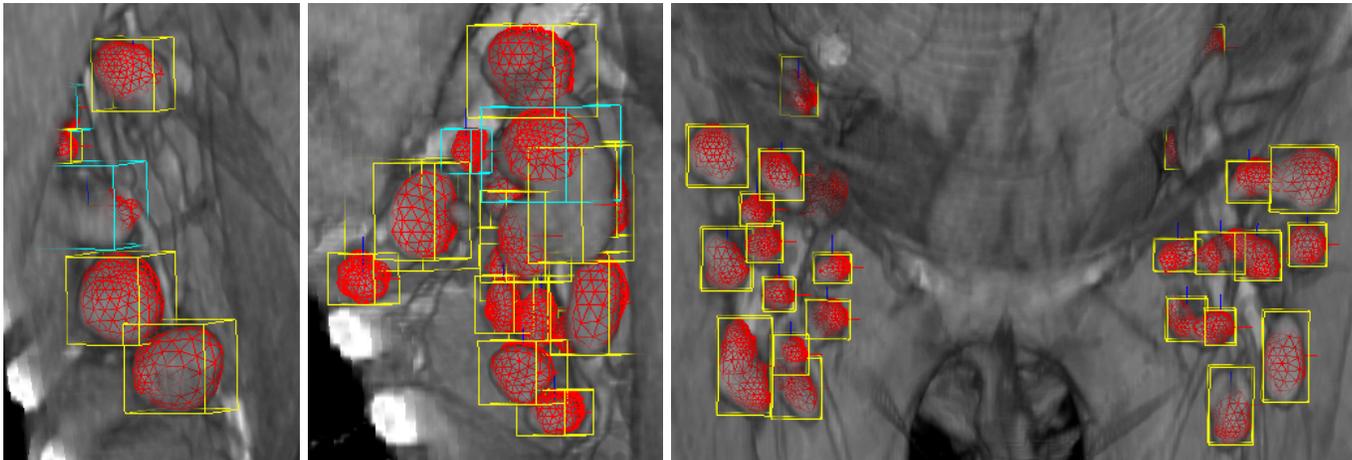| Classifier	Features	Type	# Weak Axillary	# Weak Pelvic	TPR	FPR
1	Haar	AdaBoost cascade	22, 67	25, 75	96%	$1-3\%$
2	self-aligning	AdaBoost cascade	40, 120	60, 180	94%	$2.5-6\%$
3	segmentation-based	AdaBoost	26	32	90%	$10-15\%$

Fig. 9.   Detected lymph nodes with their mesh segmentation. Middle: The method can also handle lymph node conglomerates.

A. Evaluation Datasets

Data was taken from routine CT scans of cancer patients with lymph node involvement from different hospitals in Germany, China and the United States. Intravenous contrast bolus was applied. Typical soft tissue kernels (B20 to B41) were used for reconstruction. CT slice thickness was at most 5mm.

The experiments are performed on two datasets, for axillary and pelvic lymph nodes respectively. The axillary lymph node dataset contains 131 CT volumes of 84 patients, containing 371 large ($>$ 10mm) solid lymph nodes and 546 lymph nodes that are either small ($<$ 10mm) or non-solid. The pelvic/abdominal lymph node dataset contains 54 CT volumes of 49 patients, in which the pelvic region was cropped as described in Section III-A and all lymph nodes were annotated. These pelvic regions contain 569 large solid lymph nodes and 460 small or non-solid lymph nodes. The pelvic regions contain a large number of abdominal lymph nodes, which increase the difficulty of the the detection problem as there are many folds and other structures in the abdomen that resemble lymph nodes.

All CT volumes have been converted to 1.5mm isotropic voxel size.

B. Lymph Node Annotation

All lymph nodes of size at least 10mm in the target regions of interest (axillary or pelvic+abdominal) have been annotated by placing bounding boxes around them, as shown in Figures 10 and 11. Lymph node annotation was guided by expert Radiologists who provided the major part of the annotations which were then extended by the authors. The lymph nodes are labeled as solid or non-solid depending whether they have a homogeneous interior or not.
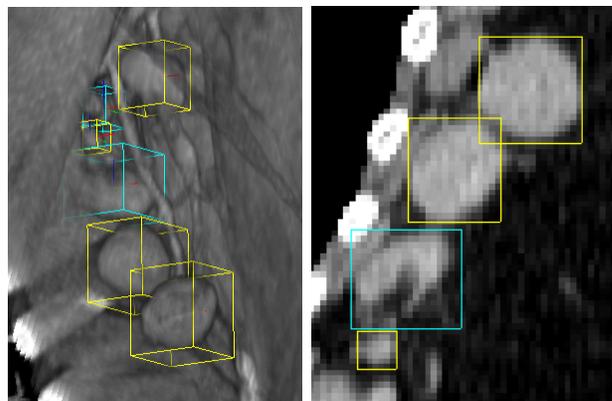

Fig. 10.   The lymph nodes are marked with bounding boxes and labeled as solid(yellow) and non-solid (cyan). Left: an axillary region. Right: a cross-section showing the difference between solid and non-solid lymph nodes.

The bounding box annotation contains all the information necessary for training the lymph node detectors described below.

C. Manual Lymph Node Segmentation

For a comprehensive evaluation we manually segmented each lymph node using an interactive segmentation tool.

The segmentation tool allows the visualization and interaction with the LN boundary on a number of cutting planes that contain the line that passes through the lymph node center and is parallel to the Z-axis. Three such cross-sections for the same lymph node are shown in Figure 12. The LN manual segmentations are also based on the sphere mesh with 162 vertices described in Section III-C. The interactive segmentation tool allows the user to manually modify the radius r of any of the 162 segmentation vertices. After a
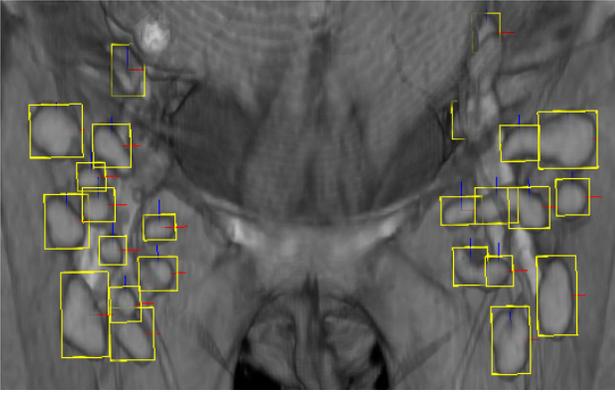
Fig. 11. A pelvic region in which the lymph nodes have been manually annotated with bounding boxes. The pelvic region usually contains both pelvic and abdominal lymph nodes.

radius r_k (having direction d_k) has been changed, the LN segmentation $\mathbf{r}$ is updated by minimizing the following energy:

$$E(\mathbf{r}) = \sum_{i=1}^{N} \alpha_i (r_i - y_i)^2 + \sum_{i=1}^{N} \frac{1}{2|\partial i|} \sum_{j \in \partial i} (r_i - r_j)^2 \quad (4)$$

where we took

$$\alpha_j = \begin{cases} 10 & \text{if } j = k \text{ or } \|d_j - d_k\| \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$
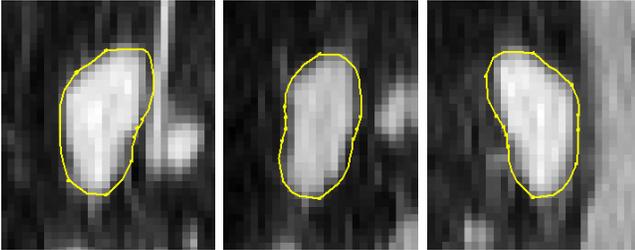

Fig. 12. The manual segmentation tool allows the LN delineation on a number of cross-sections containing the line parallel to the Z-axis that passes through the LN center.

Minimizing this energy allows the segmentation to pass very close to the modified vertex while keeping almost unchanged the vertices that are far from the interaction location. The energy minimization is done by 1000 gradient update iterations with $\eta = 0.1$:

$$r_i \leftarrow r_i - \eta \left(\alpha_i (r_i - y_i) + r_i - \frac{\sum_{j \in \partial i} r_j}{|\partial i|} \right), i = 1, ..., N \quad (5)$$

The energy minimization offers a mild smoothing of the segmentation at the vertices far away from the vertex that has been modified. If this is not desired, those vertices should not be updated at all using eq (5).

Examples of manual LN segmentations are shown in Figure 13, 14 and 15.

Figure 15 shows an elongated lymph node in 3D and two cross-sections. Using the radial representation for generating the ground truth segmentations could induce a certain bias in the generated ground truth that favors radial representations. However, we suspect that this bias is small since 97% of the lymph nodes could be accurately segmented with the manual segmentation tool described in this section.
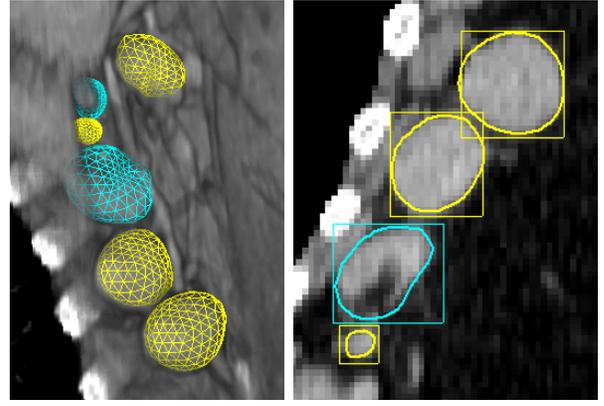

Fig. 13. The lymph nodes boundaries are delineated using a manual segmentation tool and are used to evaluate the automatic detection and segmentation.
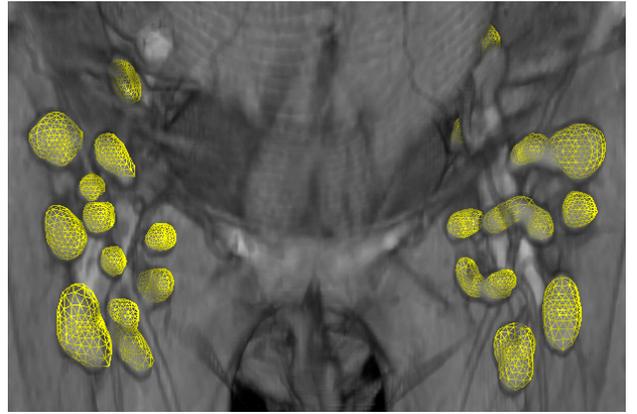

Fig. 14. A pelvic region in which the lymph nodes have been manually segmented.

D. Evaluation of Lymph Node Detection Performance

Evaluation Methodology. This approach is aimed at detecting solid lymph nodes, since they are relevant for cancer screening. Thus only the solid LN detection is evaluated. Since the solid lymph nodes are often very similar and difficult to distinguish from the non-solid ones (that have a hilum), we adopted the following evaluation measure for the detection results. A solid lymph node is considered detected if there exists a detection with the center inside the manually segmented lymph node. A detection is considered false positive if its center is not inside any annotated solid or non-solid lymph node. Thus any detection on a non-solid lymph node is neither considered a positive nor a false alarm.
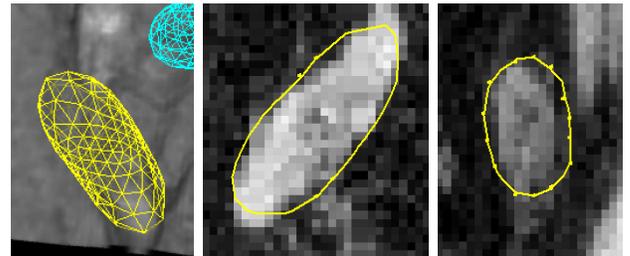

Fig. 15. A manually segmented elongated lymph node (left) and two of its cross-sections.

TABLE II
DETECTION RESULTS AND COMPARISON WITH OTHER METHODS.

| Method	Target Area	# cases	Size of LN	# LN	#TP	#FP	TPR	FP/vol	PPV	Time/vol
Ours	Axillary	131	> 10.0mm	371	308	134	83.0%	1.0	69.7%	5-20sec
Ours	Pelvic+Abdomen	54	> 10.0mm	569	455	172	80.0%	3.2	72.6%	15-40sec
Feuerstein [6] implementation	Axillary	131	> 10.0mm	371	308	4346	83.0%	33.2	0.07%	3-6min
Feuerstein [6] implementation	Pelvic+Abdomen	54	> 10.0mm	569	424	5741	74.5%	106.3	0.07%	5-10min
Feuerstein [6]	Mediastinum	5	> 1.5mm	106	87	567	82.1%	113.4	13.3%	1-6min
Kitasaka [10]	Abdomen	5	> 5.0mm	221	126	290	57.0%	58	30.3%	2-3h
Dornheim [5]	Neck	1	> 8.0mm	29	29	9	100%	9	76.3%	17min
Feulner [7]	Mediastinum	54	> 10.0mm	266	174	157	65.4%	2.9	52.6%	135sec
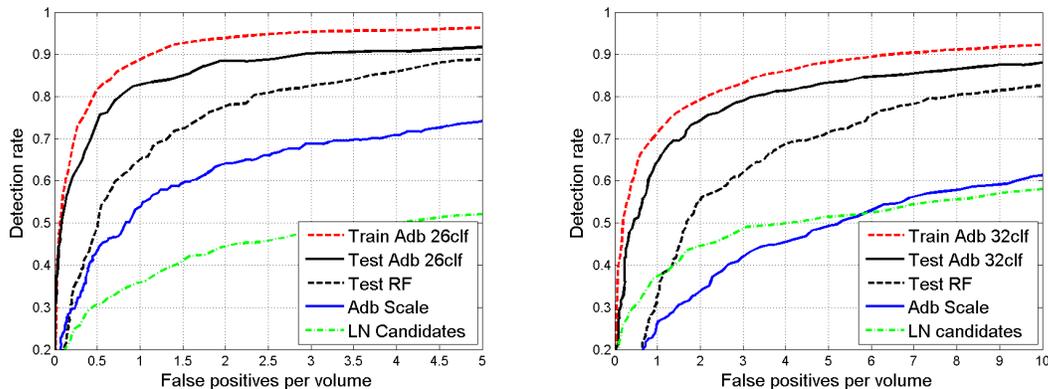
Fig. 16. Lymph node detection results with six-fold cross-validation. Left: Axillary LN evaluation on 131 volumes containing 371 LN. Right: Pelvic and abdominal LN evaluation on 54 volumes containing 569 LN.

Results.

Out of the 131 axillary cases, the region extraction failed on the left side of one patient that actually had the left lung removed. The pelvic region extraction was always successful.

Using the above evaluation measure, we obtained the ROC curves shown in Figure 16. The solid black curve represents the performance on the test set with the parameters given in Table I. The dashed red curve is the performance of the system on the training data. The dashed black curve is the performance of the system with the verification step trained as a Random Forest with 100 trees. The dash-dotted green ROC curve is the performance of the LN candidate detector. Thus, the verification step based on segmentation has a great impact on the overall performance of the system.

We also evaluated a system in which the segmentation and verification steps are removed and replaced with a lymph node detector that searches the size of the lymph node bounding box, similar to [7]. This detector is an Adaboost cascade with 50, 125 and 250 weak classifiers trained using steerable features, as described in [22]. For each of the lymph node candidates, this detector searches 890 combinations of the three lymph node sizes and reports the bounding box of highest score above a threshold. Non-maximal suppression as described in Algorithm 2 is used to further reduce the number of detections. The results using this "scale detector" are shown as a solid blue curves in Figure 16. This shows that the segmentation-based detector has a much better performance than the scale detector. An added benefit is the fact that the segmentation based detector is about 5 times faster than the scale detector. A comparison with other lymph node detection methods present in the literature is shown in Table II.

The proposed method achieves a detection rate of 83.0% at 1.0 false positives per volume for the axillary lymph nodes, i.e. a 69.7% Positive Predictive Value. On the pelvic area (that also contains many abdominal lymph nodes), the proposed method achieved a detection rate of 80.0% with 3.2 false positives per volume (72.6 positive predictive value).

These results compare favorably with the previous work [6], [10]. Dornheim[5] obtains a better detection rate but is evaluated only on a single volume.

For a more fair comparison, we implemented the approach proposed in [6] using the MinDD filter and blobness measure. For fairness, this approach was evaluated on the regions of interest (subvolumes) extracted as described in section III-A. We experimented with different parameter values and observed the best results were obtained using a blobness threshold of 5 and a size threshold of 25 for each connected component. Different values of the MinDD filter threshold gave different detection/false positive rates which are displayed as the magenta curves from Figure 17. The results of our trained detectors are shown as black curves in Figure 17 while the performance of the LN candidate detectors are shown as green dashed curves.

One could observe that the axillary dataset is easier than other LN datasets, as our implementation of [6] obtained much better results than those reported for mediastinal LN. However, the pelvic dataset appears of similar difficulty with the mediastinal LN dataset from [6]. On both axillary and pelvic datasets, our approach obtained about 30-40 times fewer false alarms for the same detection rate than our implementation of the filter based approach from [6].

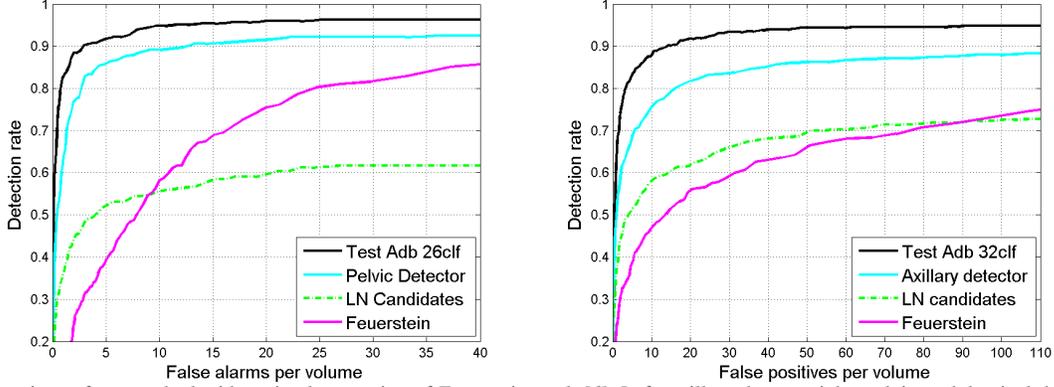To estimate the generalization ability of our trained LN

Fig. 17. Comparison of our method with an implementation of Feuerstein et al, [6]. Left: axillary dataset, right: pelvic + abdominal dataset. Also displayed is the performance of the pelvic detector on the axillary dataset (left) and the axillary detector on the pelvic+abdominal dataset (right).

detector to other types of LN, we tested the pelvic LN detector (trained on one of the six cross-validation folds) on the axillary data and the axillary lymph node detector on the pelvic data. The results are shown in cyan in Figure 17. As the pelvic detector has never seen an axillary LN, its performance is slightly below the detector trained to detect axillary LN. However, it outperforms the filter-based approach of [6] by a large margin. The same applies to the axillary detector when tested on the pelvic + abdominal dataset.

E. Evaluation of Lymph Node Segmentation Performance

The segmentation performance is evaluated only for the detected lymph nodes. We considered two evaluation measures. The first is the average symmetric point-to-mesh error, computed as the average of the distances between the automatic segmentation points and the manual segmentation mesh and vice-versa. The second is the Dice similarity coefficient (DSC) between the voxels inside the obtained segmentation and voxels inside the manual segmentation.

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

The evaluation results are summarized in Table III. Observe that the average point-to-mesh errors are smaller than the voxel size of 1.5mm. The errors are higher and the DSC is lower than the the graph based segmentation method [19], which works on CT volumes at their original resolution, with voxel size typically less than 1mm. However, it should be noted that our method is fully automatic whereas in [19] the segmentation depends on a manually given LN center.

The segmentation depends on the parameter D_{max} from eq. (1) and the data strength parameter α from eq (2). We observed experimentally that the parameter combination that gives the best detection rate ($D_{max} = 50, \alpha = 1.6$) is not the same as the parameter combination that gives the smallest segmentation error ($D_{max} = 70, \alpha = 1.6$). Hence the detection is based on a segmentation with $D_{max} = 50, \alpha = 1.6$ and for each detected lymph node another segmentation is performed with $D_{max} = 70, \alpha = 1.6$. The dependence of the point-to-mesh segmentation error on α is shown in Figure 18 (left) while the dependence on D_{max} is shown in Figure 18 (right).
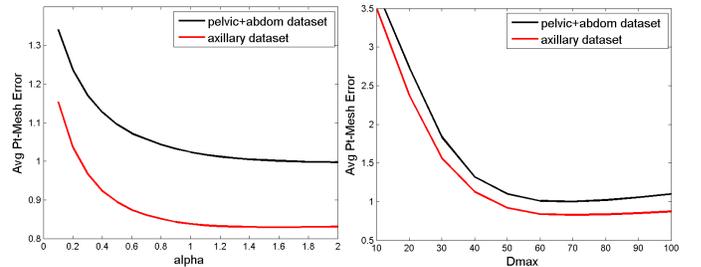

Fig. 18. Left: The symmetric pt-to-mesh error vs the parameter α of the segmentation energy (2). Right: The symmetric pt-to-mesh error vs the parameter D_{max} of (1).

As one can see, the segmentation performance is similar for a large range of values of the two parameters.
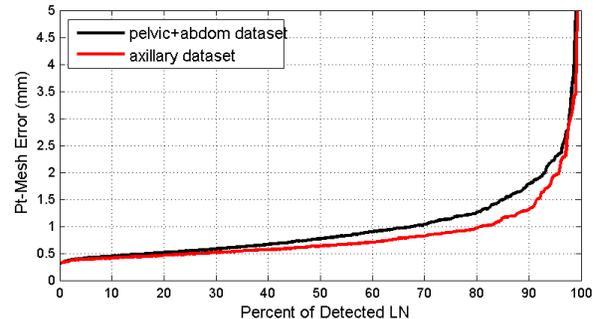

Fig. 19. The sorted point-to-mesh errors for the two datasets being evaluated.

Finally, in Figure 19 are shown the sorted point-to-mesh errors on the two datasets, which gives a better idea of the distribution of errors than the average error from Table III. From Figure 19 one could see that 80% of the axillary LN have a segmentation error less than 1mm while 80% of the pelvic+abdominal LN have an error less than 1.25mm.

Examples of segmentations for the axillary, pelvic and abdominal lymph nodes are shown in Figure 20.

V. DISCUSSION AND FUTURE WORK

In this paper, we presented a novel method for automated lymph node analysis based on integrating segmentation with a learning-based detector.

TABLE III
EVALUATION OF THE AUTOMATIC LYMPH NODE SEGMENTATION. THE STANDARD DEVIATION IS SHOWN IN PARENTHESES.

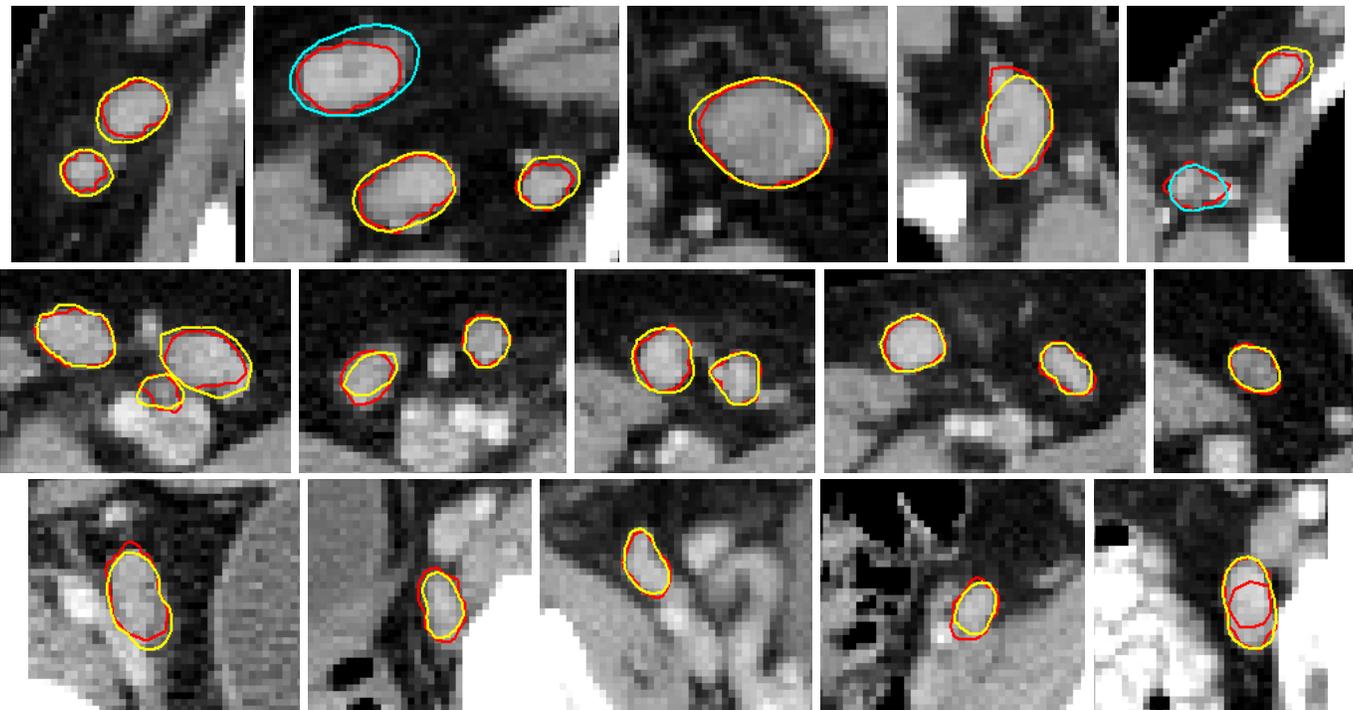| Method	Target LN	LN evaluated	Avg pt-to-mesh error (mm)	Avg DSC	Time/LN (s)
Ours	Axillary	308	0.83 (0.663)	0.80 (0.126)	0.001
Ours	Pelvic+abdom	455	1.0 (0.797)	0.76 (0.127)	0.001
Graph Based [19]	Mediastinal, abdomen, head/neck, axillary	22	0.47(0.08)	0.84	5.3
Mass-Spring Model [4]	Neck	40	0.47	-	2-30

Fig. 20. Detected and segmented lymph nodes. Top row: axillary LN, mid row: pelvic LN, bottom row: abdominal LN. Ground truth annotation is yellow for solid LN and cyan for non-solid or small LN. The segmentation result is shown in red. All results are at the same scale and the voxel size is 1.5 mm.

While we address the restricted problem of solid lymph node detection, we obtain better results with a more thorough evaluation (54-131 cases compared to 5 cases in [6], [10]). At the same time, the proposed method is also the fastest, due to detecting the lymph node centers first, while ignoring the lymph node size and shape and potentially eliminating millions of expensive verifications.

Many lymph nodes in the pelvic area are inside the abdomen and many false positives arise from the intestines and vessel bifurcations, as shown in Figure 21. A segmentation of the colon, intestines and vessels could in theory improve detection performance for the pelvic area. However, no colon or intestine segmentation has been used in the proposed approach. The axillary lymph nodes are easier to detect than other types of lymph nodes as they are far from airways and intestines, two potential sources of false positives. The evaluation confirms that the detection performance for the axillary region is better than for the pelvic area. The accuracy of the proposed approach for the axillary region could be further improved by using a vessel segmentation, to eliminate many of the systematic false positives.

In the future, we plan to study the improvement brought by using more than one segmentation at each candidate location. We also plan to use the proposed method for segmenting and detecting lymph nodes in other modalities (e.g. MRI) and other types of lesions (e.g. tumors).

REFERENCES

[1] A. Barbu, M. Suehling, X. Xu, D. Liu, S. Zhou, and D. Comaniciu. Automatic detection and segmentation of axillary lymph nodes. *MICCAI*, pages 28–36, 2010.
[2] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
[3] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, et al. Active shape models-their training and application. *CVIU*, 61(1):38–59, 1995.
[4] J. Dornheim, H. Seim, B. Preim, I. Hertel, and G. Strauss. Segmentation of Neck Lymph Nodes in CT Datasets with Stable 3D Mass-Spring Models Segmentation of Neck Lymph Nodes. *Academic Radiology*, 14(11):1389–1399, 2007.
[5] L. Dornheim and J. Dornheim. Automatische Detektion von Lymphknoten in CT-Datensätzen des Halses. In *BVM*, 2008.
[6] M. Feuerstein, D. Deguchi, T. Kitasaka, S. Iwano, K. Imaizumi, Y. Hasegawa, Y. Suenaga, and K. Mori. Automatic mediastinal lymph node detection in chest CT. In *SPIE*, volume 7260, page 30, 2009.
[7] J. Feulner, S.K. Zhou, M. Huber, J. Hornegger, D. Comaniciu, and A. Cavallaro. Lymph node detection in 3-D chest CT using a spatial prior probability. In *CVPR 2010*, pages 2926–2932.
[8] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Ann. Statist*, 28(2):337–407, 2000.
[9] A.P. Kiraly, D.P. Naidich, L. Guendel, L. Zhang, and C.L. Novak. Novel method and applications for labeling and identifying lymph nodes. In *SPIE*, 2007.
[10] T. Kitasaka, Y. Tsujimura, Y. Nakamura, K. Mori, Y. Suenaga, M. Ito, and S. Nawano. Automated extraction of lymph nodes from 3-d abdominal ct images using 3-d minimum directional difference filter. *LNCS*, 4792:336, 2007.
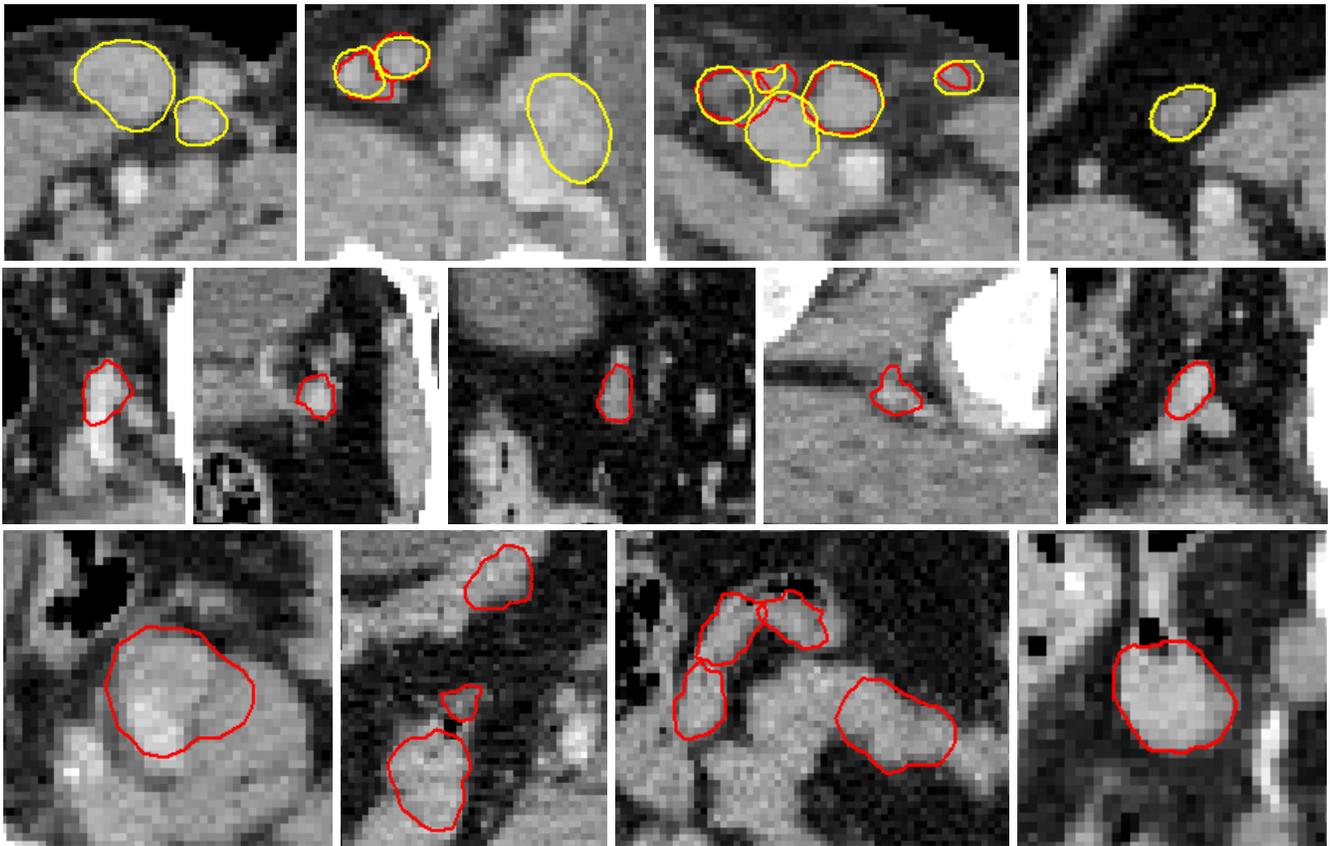
Fig. 21. Missed detections (top row) and false positives (bottom two rows) obtained with the pelvic detector. Many false positives arise on the colon, intestines and vessel bifurcations.

[11] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *IJCV*, 77:259–289, 2008.

[12] K. Li, X. Wu, D.Z. Chen, and M. Sonka. Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 119–134, 2006.

[13] R.E. Schapire. The boosting approach to machine learning: An overview. *Lect. Notes in Stat.*, pages 149–172, 2003.

[14] S. Seifert, A. Barbu, S.K. Zhou, D. Liu, J. Feulner, M. Huber, M. Suehling, A. Cavallaro, and D. Comaniciu. Hierarchical parsing and semantic navigation of full body CT data. In *SPIE Medical Imaging*, 2009.

[15] P. Therasse, S.G. Arbuck, E.A. Eisenhauer, J. Wanders, R.S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A.T. Van Oosterom, M.C. Christian, et al. New guidelines to evaluate the response to treatment in solid tumors. *Journal of the National Cancer Institute*, 92(3):205, 2000.

[16] Z. Tu, X.S. Zhou, A. Barbu, L. Bogoni, and D. Comaniciu. Probabilistic 3D polyp detection in CT images: The role of sample alignment. In *CVPR*, 2006.

[17] G. Unal, G. Slabaugh, A. Ess, A. Yezzi, T. Fang, J. Tyan, M. Requardt, R. Krieg, R. Seethamraju, M. Harisinghani, et al. Semi-automatic lymph node segmentation in LN-MRI. In *ICIP 2006*, pages 77–80.

[18] V.N. Vapnik. *The nature of statistical learning theory*. Springer, 2000.

[19] Y. Wang and R. Beichel. Graph-based segmentation of lymph nodes in ct data. *Advances in Visual Computing*, pages 312–321, 2010.

[20] J. Yan, T. Zhuang, B. Zhao, and L.H. Schwartz. Lymph node segmentation from CT images using fast marching method. *Computerized Medical Imaging and Graphics*, 28(1-2):33–38, 2004.

[21] M. Yan, Y. Lu, R. Lu, M. Requardt, T. Moeller, S. Takahashi, and J. Barentsz. Automatic detection of pelvic lymph nodes using multiple MR sequences. In *Proceedings of SPIE*, volume 6514, 2007.

[22] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE TMI*, 27(11):1668–1681, 2008.