5-2015

Segmentation of Pulmonary Nodules in Computed Tomography Using a Regression Neural Network Approach and Its Application to the Lung Image Database Consortium and Image Database Resource Initiative Dataset

Temesguen Messay
University of Dayton

Russell C. Hardie
University of Dayton

Timothy R. Tuinstra
Cedarville University, tuinstra@cedarville.edu

Recommended Citation

Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the Lung Image Database Consortium and Image Database Resource Initiative dataset

Temesguen Messay [a,*], Russell C. Hardie [a], Timothy R. Tuinstra [b]

[a] Department of Electrical and Computer Engineering, University of Dayton, 300 College Park, Dayton, OH 45469-0232, United States
[b] Department of Engineering and Computer Science, Cedarville University, 251 N. Main St. Cedarville, OH 45314, United States

ABSTRACT

We present new pulmonary nodule segmentation algorithms for computed tomography (CT). These include a fully-automated (FA) system, a semi-automated (SA) system, and a hybrid system. Like most traditional systems, the new FA system requires only a single user-supplied cue point. On the other hand, the SA system represents a new algorithm class requiring 8 user-supplied control points. This does increase the burden on the user, but we show that the resulting system is highly robust and can handle a variety of challenging cases. The proposed hybrid system starts with the FA system. If improved segmentation results are needed, the SA system is then deployed. The FA segmentation engine has 2 free parameters, and the SA system has 3. These parameters are adaptively determined for each nodule in a search process guided by a regression neural network (RNN). The RNN uses a number of features computed for each candidate segmentation. We train and test our systems using the new Lung Image Database Consortium and Image Database Resource Initiative (LIDC–IDRI) data. To the best of our knowledge, this is one of the first nodule-specific performance benchmarks using the new LIDC–IDRI dataset. We also compare the performance of the proposed methods with several previously reported results on the same data used by those other methods. Our results suggest that the proposed FA system improves upon the state-of-the-art, and the SA system offers a considerable boost over the FA system.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Lung cancer remains the leading cause of cancer death in the United States (ACS, 2013). Computed tomography (CT) is currently considered the best imaging modality for early detection and analysis of lung nodules. A wealth of image processing research has been underway in recent years developing methods for the automated detection, segmentation, and analysis of lung nodules in CT imagery (Pham et al., 2000). To facilitate such efforts, a powerful database has recently been created and is maintained by the Lung Image Database Consortium and Image Database Resource Initiative (LIDC–IDRI) (Armato et al., 2011). In this paper, we present new robust segmentation algorithms for lung nodules in CT, and we make use of the latest LIDC–IDRI dataset for training and performance analysis. Note that nodule segmentation is a critical tool in lung cancer diagnosis and for the monitoring of treatment. Multi-temporal CT scans are used to track nodule changes over certain time intervals. To make this process more accurate, consistent, and improve radiologist workflow, effective automated and semi-automated segmentation tools are highly desirable (Wormanns and Diederich, 2004). Given segmentation boundaries, nodule volume and volume doubling time can be readily computed (Ko et al., 2003; Reeves et al., 2009). For more than two decades, a variety of methods and improvements have been proposed for such lung nodule segmentation. A selected chronological listing of nodule segmentation algorithms that we believe are most closely related to our methodology is presented in Table 1. This is provided in order to put the novel contribution of our proposed methods into proper context.

While many powerful nodule segmentation methods have been proposed including the works in Table 1 and that of Coleman et al. (1998), Elmoataz et al. (2001), Wiemker and Zwartkruis (2001),

* Corresponding author.
   E-mail addresses: tmessay1@udayton.edu (T. Messay), rhardie@udayton.edu (R.C. Hardie), tuinstra@cedarville.edu (T.R. Tuinstra).

Table 1
A selected listing of nodule segmentation algorithms that are most closely related to our proposed methods.

Author(s)	Brief description
Armato et al. (1999, 2001)	Presents a complete Computer Aided Detection (CAD) where multiple gray-level thresholding and connectivity scheme are put to use to segment contiguous 3D structures. A rolling ball morphological algorithm is used to treat juxta-pleural nodules
Zhao et al. (1999a,b)	Uses multiple gray-value thresholding, 3D connected components analysis, and a 3D morphological opening operation. Features such as gradient strength and compactness are examined to determine the optimal segmentation candidate
Kostis et al. (2003)	Uses thresholding and morphological opening. Attempts to find an optimal threshold and a fixed structuring element radius suitable for all small nodules. Note that they recommend in practice the radius of the structuring element ought to be adjusted depending on the nodule under consideration. Also to note is user input is required to classify the nodule beforehand
Gurcan et al. (2004)	Patented a system that consists of, thresholding and morphological operation to get a preliminary result, adjusting the location of the supplied cue point and refining the segmentation result by using an expanded version of a fitted ellipsoid for multi-step pruning. Invention also consists of mirroring the ellipsoid about the refined cue point to create an artificially symmetric core so as to treat invasive juxtapleural nodules
Okada and Akdemir (2005), Okada et al. (2005)	Presents a scheme that makes use of an ellipsoid model using anisotropic Gaussian fitting. The volume of the nodule is estimated from the resultant ellipsoid
Kuhnigk et al. (2006)	Uses fixed thresholding followed by morphological methods. The so called *smart opening* is introduced and is adapted for each nodule. Interactive correction includes allowing the user to change the erosion strength. A convex hull operation is used to separate juxtapleural nodules
Reeves et al. (2006)	As an improvement to their earlier work in Kostis et al. (2003), an iterative algorithm that separates the nodules from the pleural surface using a clipping plane is introduced
van Ginneken (2006)	Uses a novel learning-based approach involving region growing, an iterative morphological operation, and non-linear regression. The regression system is trained voxel-wise. It uses a local 2D lung segmentation algorithm, but no evaluation for pleura-nodules is provided
Wang et al. (2007)	Uses a number of radial lines originating from the center of the volume of interest (VOI) are spirally scanned to provide a 2D projection image. Dynamic programming is then used to find the optimal outline. The favored outline is mapped to 3D space to yield the final result
Tuinstra (2008)	Makes use of multiple gray-level thresholding and morphological processing of varying strength. That engine assumes that a lung mask is provided. A trainable regression system that is similar to the one described in van Ginneken (2006) is employed to select the final nodule boundary
Li et al. (2008)	A voxel-wise segmentation approach that makes use of a 3D region growing technique is presented as part of a CAD scheme
Wang et al. (2009)	Develops a segmentation algorithm that makes use of a 3D dynamic programming model and a multi-direction fusion technique to improve the final segmentation accuracy
Moltz et al. (2009)	Presents an extension of their original work Kuhnigk et al. (2006) to improve segmentation of solid nodules located at concave parts of the pleura. An ellipsoid enclosing points obtained via ray-casting is calculated A convex hull operation, restricted to the dilated ellipsoid, is performed. The algorithm does not target non-solid nodules
Messay et al. (2010)	Makes use of a similar segmentation engine to that of Tuinstra (2008) for the CAD system. Rule based analysis and a logical operation are used to produce the final results
Kubota et al. (2011)	Proposes a voxel-wise transformation, figure-ground separation, localization of a nodule core, region growing, surface extraction and convex hull processing

Fan et al. (2002), van Ginneken et al. (2002), Xu et al. (2002), Kawata et al. (2003), Ko et al. (2003), Mullally et al. (2004), Tachibana and Kido (2006), Way et al. (2006), van Ginneken et al. (2006), Hardie et al. (2008), Kubota et al. (2008), Dehmeshki et al. (2008), Diciotti et al. (2008), Ye et al. (2009), Bendtsen et al. (2011), Gu et al. (2013), Keshani et al. (2013), Jacobs et al. (2014), none that we are aware of are able to fully and ideally address all of the challenges presented by the LIDC–IDRI dataset. Some of these advanced challenges include juxtapleural nodules that significantly invade the pleura, cases where density/intenity information is ineffectual, and non- or part-solid nodules with irregular regions-of-exclusions/cavities (Armato et al., 2011). In light of this newly expanded dataset, it behooves us to continue to explore new and more robust solutions for nodule segmentation. In this paper, we present a highly robust and novel approach for segmenting the various LIDC–IDRI nodules. Furthermore, we believe our results are among the first comprehensive nodule segmentation results produced for the new LIDC–IDRI database. Thus, it is our hope that this work may serve as a benchmark for many future nodule segmentation studies.

Our full nodule segmentation solution is a hybrid, combining a fully-automated (FA) subsystem that requires only a single centralized cue point within the nodule, and a semi-automated (SA) method that requires a set of 8 control points from the expert user. The FA subsystem builds on *the unpublished dissertation* work of Tuinstra (2008). The improvements to the method presented in Tuinstra (2008) are many fold. Note that the work of Tuinstra (2008) assumes an ideal lung mask, contoured around juxta-pleural nodules, is available *a priori*. In contrast, here we incorporate a fully automated lung segmentation algorithm. Other important advancements include, a sequence of *modified* morphological

operations that are adapted *jointly* for each nodule, a shape-model based "limiting" mechanism to treat ill-conditioned segmentation candidates and a regression neural network (RNN) that uses new salient features to evaluate the candidate segmentations. We shall show that the FA subsystem can be used alone and is competitive with other state-of-the-art systems of the same genre.

While we believe the FA system has improved robustness, some unusual and complex cases may still be problematic. Hence, if the final segmentation of the FA system is deemed inadequate the hybrid system switches to the SA subsystem. Our SA system is similar to the FA system, but uses 8 control points from the end-user. SA frameworks, that allow user intervention and/or require guiding landmarks from expert end-users for routine use in clinical settings like the ones presented in Aubin et al. (1994), Mitton et al. (2000), van Ginneken (2001), Xu et al. (2002), Pomero et al. (2004), Kuhnigk et al. (2006), Rousson et al. (2006), Dehmeshki et al. (2008), Diciotti et al. (2008), Moura et al. (2009), Moltz et al. (2009), Bendtsen et al. (2011), Vidal et al. (2011), Diepenbrock and Ropinski (2012), Gu et al. (2013) have been found to be effective in resolving advanced challenges. However, our SA system is not interactive like the nodule segmentation algorithms presented in Xu et al. (2002), Kuhnigk et al. (2006), Diciotti et al. (2008), Dehmeshki et al. (2008), Moltz et al. (2009), Bendtsen et al. (2011), Gu et al. (2013). In our proposed SA method, the required control points are entered only once. After that, the process proceeds in an automated fashion to provide the final segmentation. The extra points are used to estimate an *adaptive* shape limiting boundary that is used to impose constraints on the segmentation candidates. They are also used to modify the automatically determined lung boundary when applicable. To our best knowledge, this

particular type of user input has not been used in previously published nodule-specific studies. We think that the idea of setting the 8 points once in advance and then not manipulating the output, has the opportunity to provide more repeatability and make the radiologists workflow more consistent. Note that among other possible scenarios for incorporating extra guidance from the user, we believe that our approach effectively balances the trade-off between taxing the user and segmentation performance enhancement (i.e., a tremendous robustness and generally a large performance boost is attained in return). Also to note is that the FA system alone yields good results in many cases, and the SA system is not used.

Another novel contribution of this paper is the study of the capabilities of the FA and SA subsystems to provide segmentation characteristics that match the training truth. Since there is considerable dissent among the different radiologist truth segmentations (Armato et al., 2004, 2007, 2011), we have trained and tested multiple regression systems, each with a different form of consensus truth. This allows us to study how well our segmentation systems can adapt to various styles of truth. Finally, using our FA, SA, and full hybrid systems, we provide a thorough performance analysis and new performance benchmarks using the new LIDC–IDRI dataset.

The remainder of this paper is organized as follows. We begin by describing the LIDC–IDRI database in Section 2. The FA, SA, and hybrid nodule segmentation algorithms are described in Section 3. The RNN approach to segmentation parameter selection is described in Section 4. Experimental results and related discussion are presented in Section 5. The results include performance results on new LIDC–IDRI data, as well as a comparison with several other previously published systems using the previously available data. Finally, in Section 6 we offer conclusions. Where relevant, some of the previously published methods described in this Section are discussed further in the paper.

2. Material and methods

In this paper, we use the new LIDC–IDRI dataset (Armato et al., 2011) to train and test our algorithms. This dataset is publicly available in The Cancer Imaging Archive (TCIA), and currently contains 1010 CT scans and corresponding truth metadata (Armato et al., 2011). The truth information includes manually drawn nodule boundaries for each nodule from up to four board-certified radiologists. Details about this powerful database, such as the methods and protocols used to acquire image data, the truth annotation process, a thorough analysis of lesions, and a quality assurance evaluation, can be found in Armato et al. (2011).

2.1. The LIDC–IDRI(-) dataset

Let LIDC–IDRI(-) denote all the CT scans from LIDC–IDRI *excluding* those belonging to what used to be known as Lung Image Database Consortium (LIDC), originally hosted by National Biomedical Imaging Archive (NBIA) before the migration (Armato et al., 2004, 2007; Reeves et al., 2007; McNitt-Gray et al., 2007; Wang et al., 2007; Sahiner et al., 2007; Opfer and Wiemker, 2007; Tuinstra, 2008; Wang et al., 2009; Messay et al., 2010; Kubota et al., 2011) This LIDC–IDRI(-) subset is comprised of 926 CT scans (since the LIDC dataset contains 84 CT scans).

We have randomly selected 456 CT scans from LIDC–IDRI(-) to train and test our systems. The 456 CT scans that we use contain 432 nodules that are manually segmented by all four board-certified radiologists. We opted to use only the 432 nodules "truthed" by all four radiologists to allow us to study the impact of training and testing on various types of consensus truth. Most other

nodule-segmentation-specific studies to date (van Ginneken, 2006; Way et al., 2006; Tachibana and Kido, 2006; Wang et al., 2007; Tuinstra, 2008; Wang et al., 2009; Messay et al., 2010; Kubota et al., 2011) have used a 50% consensus criterion to combine segmentations from multiple radiologists into a single truth boundary to score their algorithm against. In this case, two or more of the four radiologists must include a given voxel in the nodule boundary to make that voxel part of the consensus truth. In addition to that common practice, here we also investigate training and testing our systems using 25%, 75% and 100% consensus truths.

To perform a rigorous validation of our systems, we have randomly partitioned the 432 nodules obtained from LIDC–IDRI(-) into three subsets, training, validation, and testing. These subsets are comprised of 300, 66, and 66 nodules, respectively. All aspects of the segmentation algorithm training and tuning is done here using the training and validation sets only (i.e., using 366 nodules). The system is then tested on the remaining 66 testing nodules. The exact testing data is publicly available through http://dx.doi.org/10.7937/K9/TCIA.2014.V7CVH1JO (download link labeled "LIDC-IDRI Image Dataset") such that it serves as an easily reproducible benchmark. Note that each expert reader has been asked to independently assess several subjective characteristics, such as subtlety, internal structure, spiculation, lobulation, shape (sphericity), solidity, margin, and likelihood of malignancy, for each lesion

Table 2
Distribution of averaged nodule characteristic ratings of the 432 nodules acquired from LIDC–IDRI(-). The ratings are on an ordinal scale of 1–5 except for calcification where the expert readers assigned a maximum rating of 6.

Subtlety rating	1	2	3	4	5	
Training data (%)	0.00	0.67	13.67	47.00	38.67	
Validation data (%)	0.00	0.00	22.73	37.88	39.39	
Testing data (%)	0.00	1.52	16.67	40.91	40.91	
Internal structure	1	2	3	4	5	
Training data (%)	99.00	1.00	0.00	0.00	0.00	
Validation data (%)	100.00	0.00	0.00	0.00	0.00	
Testing data (%)	98.48	0.00	1.52	0.00	0.00	
Calcification	1	2	3	4	5	6
Training data (%)	0.00	0.00	5.33	7.00	7.00	80.67
Validation data (%)	0.00	0.00	6.06	6.06	4.54	88.33
Testing data (%)	0.00	0.00	3.03	4.54	9.09	88.33
Sphericity	1	2	3	4	5	
Training data (%)	0.00	0.33	16.33	63.33	20.00	
Validation data (%)	0.00	1.51	25.76	59.09	13.64	
Testing data (%)	0.00	1.51	21.21	46.97	30.30	
Margin	1	2	3	4	5	
Training data (%)	0.00	1.33	8.67	42.33	47.67	
Validation data (%)	0.00	3.03	16.67	33.33	46.97	
Testing data (%)	0.00	3.03	15.15	25.76	56.06	
Lobulation	1	2	3	4	5	
Training data (%)	39.00	40.33	14.33	6.00	0.33	
Validation data (%)	33.33	48.48	13.64	4.54	0.00	
Testing data (%)	39.39	39.39	19.70	1.51	0.00	
Spiculation	1	2	3	4	5	
Training data (%)	52.67	32.33	8.67	4.67	1.67	
Validation data (%)	42.42	40.91	7.58	6.06	3.03	
Testing data (%)	48.48	40.91	7.58	3.03	0.00	
Texture	1	2	3	4	5	
Training data (%)	0.33	2.00	2.33	7.33	88.00	
Validation data (%)	1.51	0.00	1.51	7.58	89.39	
Testing data (%)	3.03	6.06	1.51	7.58	81.82	
Malignancy	1	2	3	4	5	
Training data (%)	9.67	10.00	46.67	27.33	6.33	
Validation data (%)	7.58	9.09	43.94	27.27	12.12	
Testing data (%)	6.06	10.61	46.97	33.33	3.03	

that he or she has identified as a nodule ⩾3 mm in size after the un-blinded read phase (Armato et al., 2011). Table 2 presents the distributions of the various characteristic ratings for the 432 nodules used here. The ratings are on ordinal scale of 1–5 except for calcification where the expert readers assigned a maximum rating of 6 (Armato et al., 2011; Horsthemke et al., 2010). Note that we average the individual ratings of the four readers to produce the statistics shown in Table 2. Also note that the percentage of juxtapleural nodules for the training, validation, and testing sets is 27.67%, 30.30%, and 31.82%, respectively. For a given nodule, we deduce nodule size by averaging the maximum diameter measurements in the maximum area slices. Using this method, the mean nodule sizes in the training, validation, and testing sets respectively are: 12.31 ± 5.88 mm (ranging from 4.21 to 31.62 mm); 12.96 ± 5.69 (ranging from 3.88 to 27.61 mm); and 12.88 ± 5.69 mm (ranging from 4.28 to 31 mm). Note that Table 2, and the above statistics, show we have an approximately even distribution of nodule characteristics in each of our data subsets.

2.2. The original LIDC dataset

Since many prior works on nodule segmentation have made use of the original LIDC dataset, including Wang et al. (2007, 2009), Kubota et al. (2011), we also test on this dataset to allow for a direct performance comparison. Note that since our training and validation nodules come from LIDC–IDRI(-), LIDC serves as a second independent testing set for our systems. Following the approach in Wang et al. (2007, 2009), Kubota et al. (2011) for this particular data subset, we test using only a 50% consensus truth for nodules that were segmented by three or more expert readers (out of a possible four). This leads to a total of 77 LIDC testing nodules. The original LIDC data is also publicly available via http://dx.doi.org/10.7937/K9/TCIA.2014.V7CVH1JO (download link labeled "LIDC Image Dataset") so as to aid future research efforts and comparisons. Informative works presenting details of the original LIDC dataset, such as scanner vendors, scanning protocols, reconstruction methods, and type and size of nodules, can be found in Armato et al. (2004, 2007), Reeves et al. (2007), McNitt-Gray et al. (2007), Wang et al. (2007), Sahiner et al. (2007), Opfer and Wiemker (2007), Wang et al. (2009), Tuinstra (2008), Messay et al. (2010), Kubota et al. (2011), Armato et al. (2011). After carefully examining the nodules in LIDC and LIDC–IDRI(-), it appears that the LIDC database contains a greater fraction of cavitary, irregularly-shaped, and extremely subtle nodules, compared to those of LIDC–IDRI(-). Thus, we are in agreement with Wang et al. (2009) and Kubota et al. (2011) that the LIDC dataset does indeed present difficult challenges.

3. Nodule segmentation algorithms

In this section, we describe our proposed nodule segmentation systems. We begin with the FA segmentation engine in Section 3.1. Next, we present the SA segmentation engine in Section 3.2. Finally, the hybrid system, utilizing both the FA and SA subsystems, is presented in Section 3.3. The RNN, used to automatically determine the parameters for these segmentation engines, is described in Section 4. There, both the network architecture and features are discussed.

3.1. FA method (TR segmentation engine)

A block diagram of the FA subsystem is shown in Fig. 1. The method has two free parameters, T and R. Hence, we refer to this as the TR segmentation engine. The FA subsystem assumes that

we are supplied with a CT scan in HUs, and a single well centralized cue-point in the nodule to be segmented.

As a pre-processing step, the lung fields are segmented using the automatic 3D global lung segmentation algorithm described in Messay et al. (2010). However, to improve the lung boundaries in the vicinity of juxtapleural nodules, we depart from that in Messay et al. (2010) and apply multiple successive 2D rolling ball filters of decreasing size along the outside border of the lung mask (Armato et al., 1999, 2001; Korfiatis et al., 2014). Note that all tuning parameters of the lung segmentation algorithm have been selected based on empirical studies, exclusively using the University of Texas Medical Branch data described in Ernst et al. (2004), Messay et al. (2010).

The TR segmentation engine may be viewed as a natural extension of the methods presented in Zhao et al. (1999a), Kostis et al. (2003), Gurcan et al. (2004), van Ginneken (2006), Kuhnigk et al. (2006), Tuinstra (2008), Moltz et al. (2009), Messay et al. (2010). To begin, an 80 mm³ volume of interest (VOI) around the cue point in the CT and lung mask arrays are extracted for processing. We choose the voxel that belongs to the consensus truth *and* that is closest to the centroid of the consensus truth mask to serve as the supplied cue point. This is done to define a unique cue point for each type of consensus truth discussed in Section 2. We apply the threshold T to the CT VOI data. The resulting logical array is then locally ANDed with the lung mask to exclude voxels outside the lung field and/or to disconnect juxtapleural nodules from the lung wall. The process of ANDing with the lung mask is done if and only if the delineated lung regions include the supplied user cue point. If the lung segmentation mask fails to include the cue point, which implies that the lung mask has failed to include the majority of the nodular region, we simply do not make use of the lung mask. Note that as shown in Fig. 1, if that is the case, the deployment of a limiting sphere that we are about to introduce becomes mandatory.

Next, a *modified* 2D opening that is similar to *smart* opening described in Kuhnigk et al. (2006), Moltz et al. (2009) is performed using disk-shaped structuring elements. However, we differ in that we make the dilation strength higher than the erosion. Fig. 2 shows a block diagram of the proposed modified opening. As shown in that figure, erosion is performed using a structuring element of radius R, and then dilation follows using another structuring element of radius R + 1. The modified opening is done to detach/remove residual structures, such as vessels, that may be attached to the nodule. We have discovered that using a larger dilation than erosion makes it easier to match the truth using our feature-based regression system. This is likely due to the fact that the provided truth outlines are intended as "outer borders" that do not overlap with voxels belonging to the nodule (Armato et al., 2011). Features such as intensity gradients tend to favor smaller segmentations where the intensity inflection boundary occurs. Note that the proposed modified opening requires the specification of the parameter R only. The static +1 offset has been selected based on empirical study, exclusively using the 300 training nodules. The study shows +1 to be the best offset on average with respect to the multiple types of consensus truth. The area criterion (less than 2 mm²) shown in Fig. 2, used to remove small structures prior to dilating, has been similarly determined using the training data. We recommend 2D morphological processing here (i.e., in each cross-sectional CT slice) because of the non-isotropic nature of the LIDC–IDRI data (Armato et al., 2011).

After the modified opening, we enforce connectivity to the cue point, as shown in Fig. 1. In some difficult cases, the segmentations at this point in the system may still include significant non-nodular anatomical structures. It is for that reason that we include the limiting sphere block in Fig. 1. Here, if the candidate segmentation exceeds a size threshold, or if the cue point is outside the
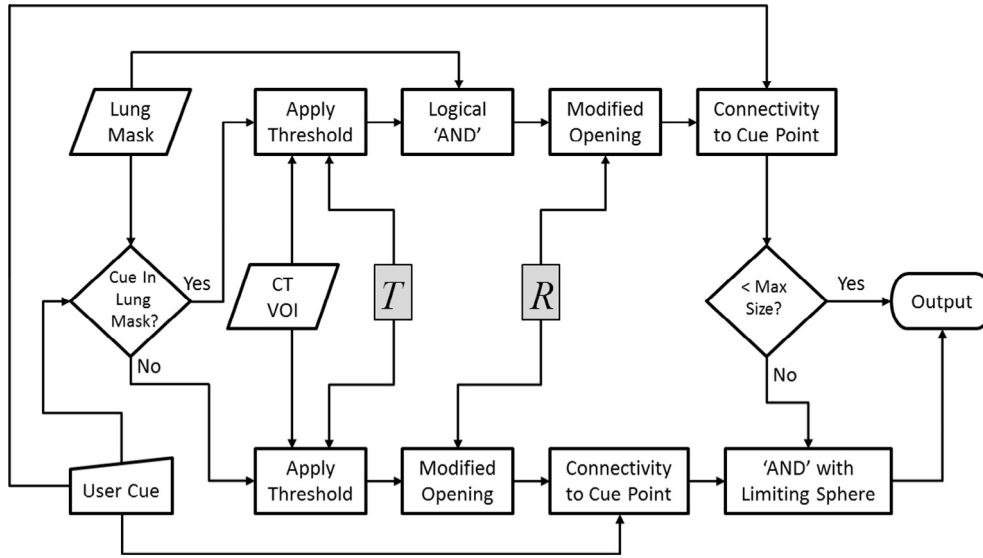
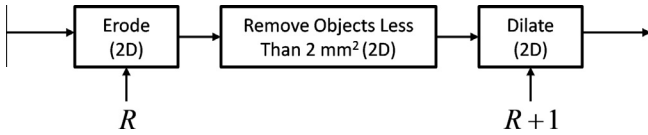**Fig. 1.** FA subsystem (TR segmentation engine) block diagram.

Fig. 2. Modified 2D opening operation.

computed lung mask, the candidate mask is logically ANDed with a sphere obtained using the ray shooting/casting technique described in Moltz et al. (2009). However, instead of using a fixed threshold of −400 HU, as is done in Moltz et al. (2009), we use a threshold that is slightly lower than the intensity/density at the supplied cue point. The median of the "valid" radial distances (Moltz et al., 2009) is multiplied by 1.15 to give the radius of the limiting sphere. We believe that this mechanism is a practical way to add robustness without greatly increasing computational complexity. Note that although the described provision was originally intended to address unusual cases, we have found that it is effective in many well-conditioned cases as well. The limiting sphere takes some of the "burden" of pruning non-nodulate structure off of the opening operation. This often allows for the use of smaller R values, which tend to better preserve detail on the nodule boundary (Serra, 1983; Strickland, 2002; van Ginneken, 2006). We believe that this relatively simple segmentation engine can be exceptionally powerful, provided that the T and R parameters are jointly optimized for each nodule. Of course, the big challenge lies in how to best determine these parameters automatically. This is addressed in Section 4.
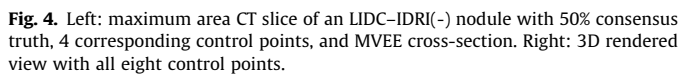
3.2. SA method ($\widetilde{T}RE$ segmentation engine)

The SA subsystem block diagram is shown in Fig. 3. The free parameters, which will be explained below, are the threshold $\widetilde{T}$, structuring element size R, and ellipsoid scaling parameter E. As such, this method will be designated the $\widetilde{T}RE$ segmentation engine. In addition to the CT scan and lung mask, the $\widetilde{T}RE$ segmentation engine requires eight control points from the expert end-user. These eight control points should include the four end-points of the major and minor axes of the nodule in the maximum nodule area slice. The end points of the major axes of the first and last nodule-containing slice make up the other 4 points. Note that major

and minor axes here are the length and width of the nodule as defined by the ELCAP protocol (Henschke et al., 2002; Kubota et al., 2011). In this work, unique sets of control points are extracted from the multiple types of consensus truth. From the 8 user points, we compute the minimum volume enclosing ellipsoid (MVEE) (Gurcan et al., 2004; Okada et al., 2005; Okada and Akdemir, 2005; Moshtagh, 2005; Moltz et al., 2009). The MVEE is scaled in size by the scalar parameter E and used to bound segmentation mask. For nodules that only appear in one CT slice, the user is only required to provide the four control points from the maximum area slice. In that case, a 2D enclosing ellipse is computed and a scaled version is used to bound the segmentation. Fig. 4 illustrates the 8 control-points and fitted shape model for an LIDC–IDRI(-) nodule. Note that the scaled ellipsoid is truncated in the axial direction, so as to not extend past the user points specifying the top and bottom slices.

We acknowledge that this novel 8 cue points approach is more demanding on the user, compared to the FA subsystem and other single cue point systems (Kostis et al., 2003; Gurcan et al., 2004; van Ginneken, 2006; Wang et al., 2007; Tuinstra, 2008; Wang et al., 2009; Moltz et al., 2009; Kubota et al., 2011). Be that as it may, the open-minded reader will recognize that this is far less burdensome than manually delineating the boundary of the nodule in its entirety. Moreover, we believe that our approach may lead to more consistent segmentation results by using the well defined 8 control points designation task, and potentially eliminating the need for subjective post-segmentation mask editing. This could potentially improve the accuracy of volume doubling time analyses. Furthermore, SA is only utilized in our overall hybrid system when necessary (i.e., treating special/irregular cases that present difficult challenges). To make the acquisition of these 8 points as fast and convenient as possible, a special graphical user interface can be employed to allow the user to "drag and drop" the linear extents of all four axes, while enforcing the orthogonality of the two axes at the max area slice. Among other possible scenarios for incorporating extra guidance from the user, we believe that our approach effectively balances the trade-off between taxing the user and segmentation performance enhancement.

The basic functionality of the SA subsystem is similar to that of the FA subsystem. It is based on thresholding, opening, bounding, and connectivity. However, there are some important differences in other aspects of the system. First of all, for the SA system, the

Fig. 3. SA subsystem ($\widetilde{T}RE$ segmentation engine) block diagram.

Fig. 4. Left: maximum area CT slice of an LIDC–IDRI(-) nodule with 50% consensus truth, 4 corresponding control points, and MVEE cross-section. Right: 3D rendered view with all eight control points.

VOI extracted from the CT data and lung mask is not of fixed size. Rather, the VOI size is selected so as to contain the 8 points with appropriate padding, so as to facilitate feature computation. Also, unlike the *TR* segmentation engine, we do not directly set the threshold as an input in HUs. Instead, the threshold is based on the mean and standard deviation of the voxels along the major and minor axes at the max area slice, and along the major axes at the first and last slices. In particular, the applied threshold is given by $T = \mu - \widetilde{T} \times \sigma$, where $\widetilde{T}$ is the input tuning parameter and μ and σ are the mean and standard deviation in HUs, respectively.

Another important distinction of the $\widetilde{T}RE$ segmentation engine is the revised lung mask concept. Note that in Fig. 3, the MVEE is logically ORed with the computed lung mask. This guarantees that the MVEE is considered part of the lung and yields what we refer to as the "revised" lung mask. This revised lung mask allows us to cope with juxtapleural nodules that significantly invade the pleural surface. Three such nodules from the LIDC–IDRI(-) datatset are shown in Fig. 5. Our standard lung masks are shown in green on the left in Fig. 5. The revised lung masks, that incorporate the MVEE, are shown on the right. The red contours are the 50% consensus truth masks. The rolling ball technique in Armato et al. (1999, 2001), Messay et al. (2010), Korfiatis et al. (2014) is effective at compensating for the indentations along the contour lines of each lung caused by juxtapleural nodules. However, it is insufficient for the purpose of segregating the perimeter of pleura-nodules that are significantly embedded in the lung wall. The pleural segmentation algorithm (clipping plane), proposed by Reeves et al. (2006) to approximate the pleural surface, is also inadequate in resolving the problem. It only works well when the surface between the nodule and the lung wall can be approximated by a plane (Reeves et al., 2006). The convex hull method, described in Kuhnigk et al. (2006), has similar difficulty with embedded juxtapleural nodules, since it implicitly assumes that the average boundary of the lung is smooth. A ray casting approach is suggested in Moltz et al. (2009) to segment nodules that are attached to non-convex parts of the pleura. Their latest suggestion assumes that the points found by the ray casting procedure cover a major part of the actual nodule surface (Moltz et al., 2009). That is not always the case for some LIDC–IDRI nodules. For example, consider the juxtapleural nodule depicted in Fig. 5(e) and (f). Performing the region growing and ray casting procedure on this nodule will yield few valid ray end-points, and these end-points will not fully capture the geometry of the nodule. The mirroring method, described in Gurcan et al. (2004), may also be inadequate to resolve the challenge presented by this particular nodule due to the severe asymmetry with respect to the original segmented lung boundary. For these reasons, we believe that using the MVEE from the extra user points to revise the lung mask to be an effective and robust solution.

The final processing block for $\widetilde{T}RE$ segmentation engine shown in Fig. 3, enforces a connectivity constraint. Here we impose a 6-connected 3D connectivity rule to the 8 user points, plus a computed centralized point (9 points total). The "central point" is computed by finding the intersection of the supplied major and minor axes in the max area slice.

Note that the three free parameters, $\widetilde{T}, R$, and E, are to be selected jointly for each nodule in an adaptive fashion. These parameters can work together in interesting ways. The scaled MVEE bounding method can be thought of as another way to remove attached non-nodule structures, but in a much more targeted manner than we are able to do in the *TR* engine. This relieves the opening operation of the sole duty of pruning the thresholded segmentation. Also, since the bounding ellipsoid is a scaled version of the MVEE, the system can be tuned to provide just the necessary pruning, or even no pruning at all. This is particularly important for non-elliptical nodules. At the same time, we are still providing a natural nodule boundary, that is based on thresholding and gentle opening, for the vast majority of the nodule surface. This is in contrast to Okada and Akdemir (2005), Okada et al. (2005), where the volume of the nodule is obtained using a fitted ellipsoid itself. One other powerful benefit of the user input based MVEE, is that it reflects the desired shape parameters sought by the expert user.
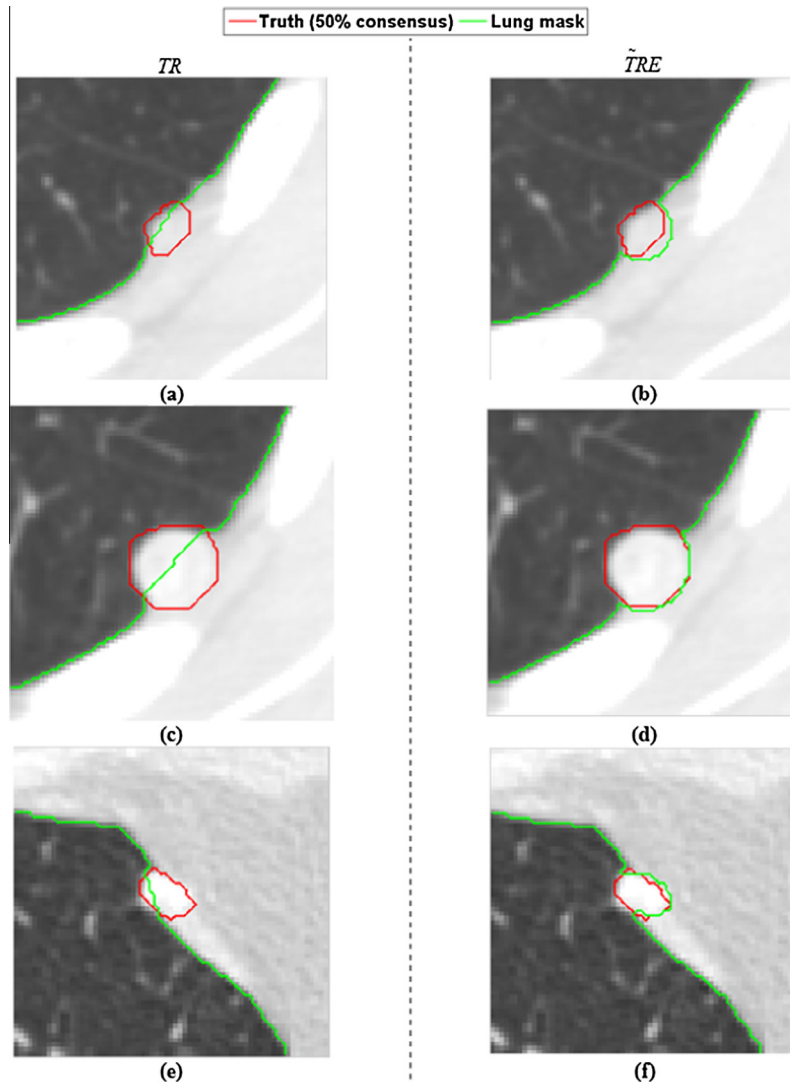
Fig. 5. Three nodules from LIDC–IDRI(-) with corresponding 50% consensus truth (red) and lung masks (green). Left: standard lung mask used by the TR engine. Right: revised lung mask formed by ORing with the MVEE in the $\widetilde{T}RE$ engine. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This way the system is better able to deliver a final segmentation in accordance with the desires of the user. In Section 4, we address how the parameters of $\widetilde{T}RE$ subsystem can be found jointly in an automated manner.

3.3. Hybrid segmentation system

The full proposed nodule segmentation system is a hybrid that combines the FA and SA segmentation engines. Fig. 6 shows a top level block diagram of this hybrid system. To start with, the system requires a single central cue point to initiate the TR segmentation engine of the FA subsystem. The result is analyzed to determine if the TR segmentation is adequate. If it is deemed to be adequate, the resulting mask may be used as the final output and processing is complete. However, if the TR segmentation is deemed to be inadequate (or there is a desire to seek improvement), the SA subsystem is launched and the user is cued to provide the 8 control points required. The decision rule can be a manual, one controlled by the user, or an automated one built into the system. For our automated approach, we employ a relatively simple decision rule. This decision rule is based on the estimated overlap score (EOS) for each output segmentation as provided by the RNN described in

Section 4. The true overlap score is commonly used as a segmentation performance measure and it is defined to be the size of the intersection set of the true and estimated segmentation masks divided by size of the union set (van Ginneken, 2006; Wang et al., 2007; Tuinstra, 2008; Wang et al., 2009; Messay et al., 2010; Kubota et al., 2011). EOS can be used as a measure of confidence in the TR segmentation. The other factor that we find to be highly relevant, is the amount of contact a nodule candidate has to the pleural surface. Since the TR segmentation engine is known to have limitations for deeply embedded juxtapleural nodules, we have decreasing confidence as the amount of pleural contact increases. Thus, our automated system declares a TR segmentation to be adequate if the EOS is greater than 70% *and* the fraction of segmented voxels along the outer boundary of the mask that are in contact with the lung wall is less then 0.3. Otherwise, the system will recommend to the user that the SA module, using the $\widetilde{T}RE$ engine, be launched. Note that it is possible to provide the user with multiple TR candidate segmentations, based on training with different truth, for consideration at this stage. If the $\widetilde{T}RE$ segmentation engine is launched, the user is able to significantly control the resulting segmentation mask through the 8 provided cue points. If the system is trained on multiple
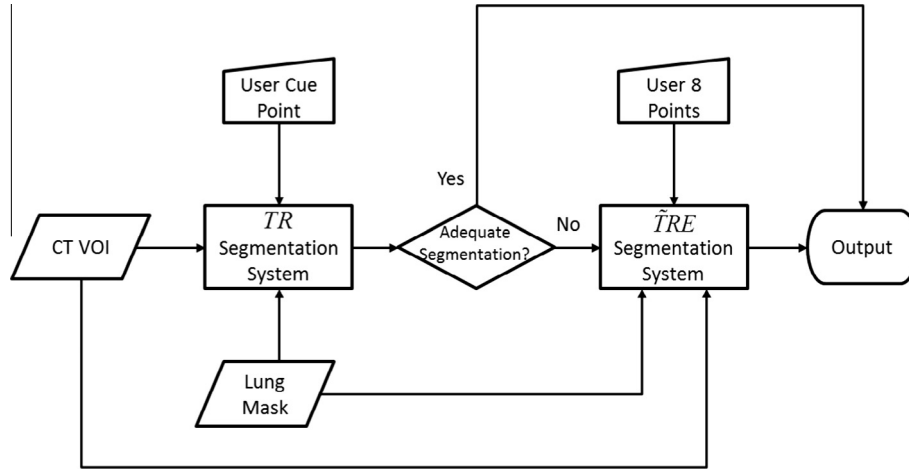
Fig. 6. Hybrid nodule segmentation system combining the FA and SA segmentation engines.

truths, multiple $\widetilde{T}RE$ outputs could be presented to the user for final selection. Note that the user may also choose to reposition the control points, as deemed fit, and rerun the $\widetilde{T}RE$ segmentation engine, or restart the hybrid system from the start.

4. Segmentation engine parameter selection

4.1. Regression neural network

Given the parametric segmentation engines, the challenge is to automatically, and jointly, optimize the engine parameters for each nodule. Of course, such parameters could be set manually, as mentioned in Tuinstra (2008). To automate the process, our system trains and deploys an RNN to serve as a computational/artificial expert to "grade" segmentations. The RNN takes as input an extensive set of salient features, computed from the CT data and the candidate segmentation mask, and produces an EOS. Note that feature-based trainable algorithms have been previously proposed by van Ginneken (2006) and Tuinstra (2008). This approach is attractive in the sense that it can not only be applied to different types of nodules, but also adapt to the multiple types of truth by simply changing the training data. Other methods, like the ones described in Zhao et al. (1999a), Gurcan et al. (2004), van Ginneken (2006), Hardie et al. (2008), Tuinstra (2008), Messay et al. (2010), that optimize segmentation parameters based on computed features have also been developed. However, most of these prior methods have used a small number of features and a simple rule based parameter selection. To the best of our knowledge, here we use the most extensive set of features applied to this problem to date. More will be said about the features and feature selection in Section 4.2. Because of the large number of features, a simple rule based parameter selection is not feasible. Thus, we employ an RNN to process the features and embody the knowledge-base contained within the training set. The availability of the expanded dataset in LIDC–IDRI makes the use of an expanded feature set and RNN feasible (Wang et al., 2000; van Ginneken, 2006).

Fig. 7 shows how the RNN is used to evaluate candidate segmentations and allow us to optimize the segmentation engine parameters in a custom manner for each nodule. Note that this same approach is used for the TR and $\widetilde{T}RE$ segmentation engines. After a candidate segmentation is generated, a set of features is computed and fed into the neural network. The resulting EOS is then fed into an adaptive algorithm that controls the segmentation
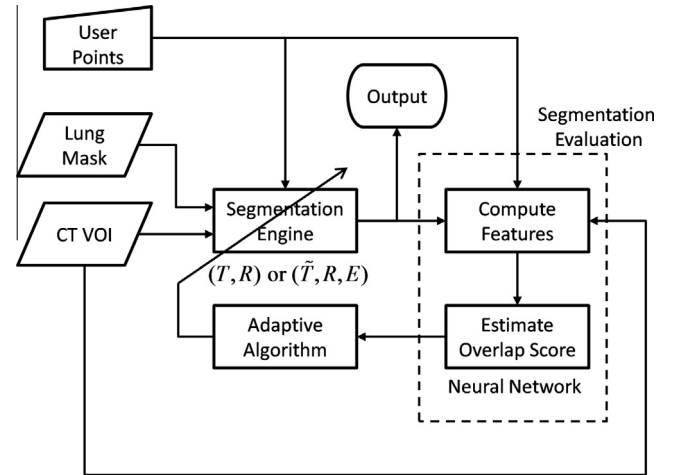

Fig. 7. Regression neural network based nodule segmentation evaluation and engine parameter selection.

engine tuning parameters. The goal is to find the candidate that produces the highest EOS, and this candidate is selected as the output for the segmentation engine. In the results presented in this paper, we have elected to use an exhaustive search over a fine grid of parameters. The engine parameters that give the highest EOS score over a fine grid of parameters are selected for the final segmentation. The idea is to focus on top performance, and not processing speed, at this stage in the research. Future work may focus on acceleration of the algorithm implementation.

To create the necessary training data, we generate candidate segmentations for all of the training nodules using a uniform grid of segmentation engine parameters. For the TR system, we use $R \in \{0, 1, 2, 3\}$ and $T \in \{-1024, -1020, -1016, \ldots, -4\}$. This generates a total of 1024 different segmentation candidates for each nodule. The cue point for training is automatically generated by using the voxel within the truth mask that is closest to the centroid. For $\widetilde{T}RE$, we let $\widetilde{T}$ take on 65 linearly spaced values ranging from 0 to 4. For the parameter E (normalized ellipsoid scaling values), we use 8 values ranging from 1.01 to 2.5. Note that the bounding ellipsoid is required to be larger than the MVEE during scaling, so as to avoid cropping one of the user specified cue points. The parameter R in $\widetilde{T}RE$ serves the same purpose as in TR, hence it takes on the same range of values. This results in a total of 2080

segmentation candidates for every nodule.. After the segmentation candidates are created, features and overlap scores are computed for each segmentation candidate. These are used for RNN training. Since we are considering two segmentation engines and 4 types of consensus truth, we train a total of 8 RNNs.

In our approach, we make use of Multi-Layer Perceptron (MLP) RNNs to produce an EOS for each candidate (Rogers, 1991; Kosko, 1992; Girosi et al., 1995; Ham and Kostanic, 2000; Wang et al., 2000; Tuinstra, 2008). We have chosen an MLP-RNN with one hidden layer, one bias node at each level, and initially 40 hidden nodes. The output layer consists of a single node plus a bias term, since the EOS is the only output. The network's input neurons are fed by the computed features that are scaled to lie between -1 and 1 prior to training and are associated with input bias nodes. We make use of hyperbolic tangent sigmoidal functions in the hidden layer, and a logarithmic sigmoidal function in the output layer to constrain the output to lie between 0 and 1 (Rogers, 1991; Kosko, 1992; Girosi et al., 1995; Ham and Kostanic, 2000; Wang et al., 2000; Tuinstra, 2008). The weights and biases are determined during network training. Initially, weights and bias values are adjusted using the training data until the error reaches a plateau. At that stage, the validation data are employed. If the validation performance fails to decrease for six successive epochs, the training is terminated (Rogers, 1991; Kosko, 1992; Girosi et al., 1995; Ham and Kostanic, 2000; Wang et al., 2000). The entire process described above (including the validation checks) is repeated for different numbers of neurons and different numbers of features. The idea is to create an RNN that is able to generalize well, while maintaining a good overall performance. All eight RNNs are trained separably using the training and validation data-sets (300 and 66 nodules, respectively). The final architectures for each system are provided in Table 3.

4.2. Features

We now turn our attention to the features used by the RNNs. A proper choice of features is important to obtain good performance and generalizability (Rogers, 1991; Kosko, 1992; Girosi et al., 1995; Ham and Kostanic, 2000; van Ginneken, 2006; Tuinstra, 2008). We have experimented with various feature selection methods, including the reaction based approach described in Verikas and Bacauskiene (2002). Here, our features are selected using a combination of methods. First, we include intuitively justifiable geometric and segmentation parameters. The remaining features are selected from the large pool of intensity and gradient features described in Hardie et al. (2008), Tuinstra (2008), Messay et al. (2010) using a correlation analysis. Table 4 shows the final list of selected features for the four *TR* RNNs. In that table, $TR25, TR50, TR75$ and $TR100$ denote RNNs optimized for $25\%, 50\%, 75\%$ and 100% consensus truth, respectively. Table 5 shows a similar list of selected features for the four $\widetilde{T}RE$ RNNs. The only difference between the list of possible features in Tables

Table 3
MLP-RNN architectures used for segmentation parameter selection.

System	Truth (%)	Features	Hidden nodes
*TR*25	25	29	20
*TR*50	50	28	30
*TR*75	75	28	30
*TR*100	100	28	30
$\widetilde{T}RE$25	25	31	30
$\widetilde{T}RE$50	50	31	30
$\widetilde{T}RE$75	75	31	30
$\widetilde{T}RE$100	100	30	10

Table 4
List of selected features for the four *TR* segmentation systems (each optimized with respect to a unique type of consensus truth). The features are computed using the boundary defined by the segmentation mask and the CT data in HU. 2-D features are computed at the maximum area slice.

	Systems			
	TR25	TR50	TR75	TR100
2-D geometric features				
Size (Major Axis Length)	×	×	×	×
Circularity	×	×	×	×
3-D geometric features				
Volume	×	×	×	×
Sphericity	×	×	×	×
Elongation	×	×	×	×
Juxtapleural	×	×	×	×
Fraction Touching Lung	×	×	×	×
2-D intensity features				
Contrast	×			
3-D intensity features				
Contrast		×	×	×
Standard Deviation Separation	×	×	×	×
2-D gradient features				
Radial-Deviation Mean Outside		×	×	×
Radial-Deviation Mean Separation	×	×	×	×
Radial-Deviation Standard Deviation Inside	×	×	×	×
Radial-Deviation Standard Deviation Outside	×	×	×	×
Radial-Gradient Mean Outside	×	×	×	×
Radial-Gradient Mean Contrast	×			
Radial-Gradient Standard Deviation Inside	×	×	×	×
Radial-Gradient Standard Deviation Separation	×	×	×	×
3-D gradient features				
Radial-Deviation Mean Inside		×	×	×
Radial-Deviation Mean Outside	×		×	×
Radial-Deviation Mean Contrast	×	×		
Radial-Deviation Standard Deviation Inside	×			
Radial-Deviation Standard Deviation Outside	×	×	×	×
Radial-Deviation Standard Deviation Separation		×	×	×
Radial-Gradient Mean Inside	×	×	×	×
Radial-Gradient Standard Deviation Inside	×	×	×	×
Radial-Gradient Standard Deviation Outside	×			
Radial-Gradient Standard Deviation Separation	×	×		×
Segmentation parameter based features				
Sphere radius	×	×	×	×
Mask centroid error	×	×	×	×
T	×	×	×	×
R	×	×	×	×
T at cue point	×	×	×	×

4 and 5, is the segmentation parameter based features at the bottom of these table.

Let us briefly describe the key features, starting with those in Table 4 for the *TR* segmentation engine. The hand-picked features include 2D and 3D geometric features that are described in Giger et al. (1988, 1990, 1994), Armato et al. (1999, 2001), Hardie et al. (2008), Messay et al. (2010). We also include the segmentation engine parameters used to generate the candidate mask in question. For the *TR* engine, this includes T, R, and the bounding sphere radius. In the case where the sphere is not employed, we set the sphere radius feature to 80 mm (the max dimension of the VOI). The feature denoted as *Mask Centroid Error*, is the Euclidean distance between the coordinates of the supplied seed point and the computed centroid of the segmentation candidate. The feature *T at cue point*, refers to the density in HU at the supplied cue point. The remaining selected features in Table 4 come from the pool of features described in Hardie et al. (2008), Tuinstra (2008), Messay et al. (2010). The bulk of this feature pool is made up of features that have been found to be useful for nodule detection in our

Table 5
List of selected features for the four $\widetilde{T}RE$ segmentation systems.

	Systems			
	$\widetilde{T}RE25$	$\widetilde{T}RE50$	$\widetilde{T}RE75$	$\widetilde{T}RE100$
2-D geometric features				
Size (Major Axis Length)	×	×	×	×
Circularity	×	×	×	×
3-D geometric features				
Volume	×	×	×	×
Sphericity	×	×	×	×
Elongation	×	×	×	×
Juxtapleural	×	×	×	×
Fraction Touching Lung	×	×	×	×
2-D Intensity features				
Standard Deviation Separation		×	×	
3-D Intensity features				
Contrast	×	×	×	×
Standard Deviation Separation	×			×
2-D gradient features				
Radial-Deviation Mean Inside	×	×	×	×
Radial-Deviation Mean Outside	×	×	×	
Radial-Deviation Standard Deviation Inside	×	×	×	
Radial-Deviation Standard Deviation Outside	×	×		
Radial-Deviation Standard Deviation Separation				×
Radial-Gradient Mean Inside		×	×	×
Radial-Gradient Standard Deviation Inside				×
3-D gradient features				
Radial-Deviation Mean Inside	×	×	×	
Radial-Deviation Mean Outside		×	×	×
Radial-Deviation Mean Contrast	×			×
Radial-Deviation Standard Deviation Inside	×			×
Radial-Deviation Standard Deviation Outside		×	×	
Radial-Gradient Mean Inside	×	×	×	
Radial-Gradient Mean Outside		×	×	×
Radial-Gradient Mean Contrast	×			
Radial-Gradient Standard Deviation Inside	×		×	×
Radial-Gradient Standard Deviation Outside				
Radial-Gradient Standard Deviation Separation	×	×	×	×
Segmentation parameter based features				
$\widetilde{T}$	×	×	×	×
R	×	×	×	×
E	×	×	×	×
2-D 4-Points Match Mean Error	×	×	×	×
2-D 4-Points Match Error Variance	×	×	×	×
3-D 8-Points Match Mean Error	×	×	×	×
3-D 8-Points Match Error Variance	×	×	×	×
3-D Number of Slices Fractional Error	×	×	×	×

CAD system (Messay et al., 2010), but most have not previously been used for nodule segmentation.

To determine the most salient intensity and gradient features from the feature pool, we first compute the correlation coefficient (Papoulis and Pillai, 2002; Lay, 2003) between the features and the computed overlap scores for the corresponding nodules. Note that we first slightly erode the segmentation candidate mask before computing those features. Fig. 8 illustrates the magnitude of the gradient field for one of the training nodules. As it can be seen in that figure, the overlaid multiple consensus truth are slightly larger relative to the boundary of the nodule that is highlighted by the maximum strength of the gradient field. This erosion process helps align the maximum overlap score with the peak in most gradient based features. Since our segmentation engines perform the modified opening, which expands our final segmentation mask, we believe such erosion. during feature computation is appropriate. The correlation based feature selection procedure continues as follows. The feature from the pool with the highest correlation coefficient is appended to the list of potential salient features. Remaining features in the pool, with a correlation coefficient magnitude ⩾0.8 with respect to the currently selected feature, are discarded from future consideration. We repeat this process until we get to 65 features. At this point, any highly linearly dependent features are discarded (Papoulis and Pillai, 2002; Lay, 2003). The final phase of feature selection is based on maximizing system performance on the validation data.

The same process is used to obtain the $\widetilde{T}RE$ features shown in Table 5. The only difference here is that we have computed and include additional features that make use of the extra user-supplied control points. These extra features include the mean and variance of the Euclidean distances between the supplied control points and the nearest corresponding points on the surface/perimeter of the candidate segmentation. The feature in Table 5 named *3-D Number of Slices Fractional Error* is the ratio of the number of slices between the top and bottom user cue points and the number of slices in the candidate segmentation. These extra features are powerful and allow us to evaluate how well a candidate segmentation conforms to the user provided shape information.

5. Experimental results and discussion

In this section, we present several results to demonstrate the efficacy of the proposed systems. First, we present quantitative performance results for TR, $\widetilde{T}RE$, and the hybrid system using our LIDC–IDRI(-) testing data. Next, we present results using LIDC data, so that we may directly compare the performance of the proposed systems to several other reported segmentation systems using the same data. This section concludes with a discussion of the systems and results.

5.1. Performance on LIDC–IDRI(-)

A performance analysis summary for the segmentation systems proposed here, using the LIDC–IDRI(-) dataset, is provided in Table 6. Results are reported for the FA system (TR engine), the SA system ($\widetilde{T}RE$ engine), and the automated hybrid system. We have trained the systems using the training and validation data with 25%, 50%, 75%, and 100% consensus truths. The overlap scores presented are relative to the type of truth used for training. These overlap scores are reported for all three LIDC–IDRI(-) data subsets (training, validation and testing). The cue point used for the TR system is set as voxel that lies within the specified truth mask that is closest to the centroid of the truth mask. For the $\widetilde{T}RE$, the 8 control points are obtained as described in Section 3.2 from the truth mask.

The results on the testing data are the primary results. The results on training and validation are for diagnostic purposes only. The columns in Table 6 labeled "sphere cases" refer to the number of nodules where the TR engine deploys the limiting sphere in producing its output segmentation, as described in Section 3.1. The column labeled "TRE cases" refers to the number of cases where the automated hybrid system determined that the TR engine segmentation was inadequate, and deployed the $\widetilde{T}RE$ engine, with its extra cue points. We have highlighted the 50% consensus truth portion of the table, because this is the most common method used in the literature (Tachibana and Kido, 2006; Way et al., 2006; van Ginneken, 2006; Wang et al., 2007; Tuinstra, 2008; Wang et al., 2009; Messay et al., 2010; Kubota et al., 2011). Note that the TR
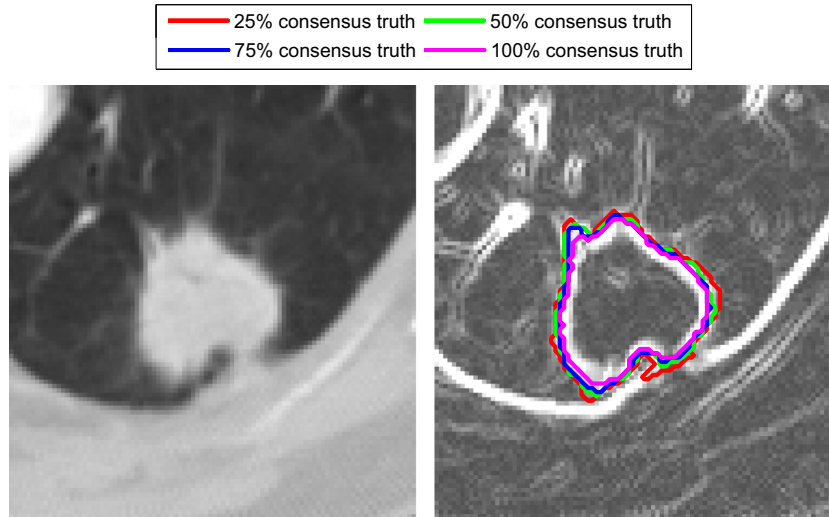
Fig. 8. Shown to the left is a nodule from our training data. On the right is the magnitude gradient field overlaid with the various types of consensus truth.

Table 6
Performance analysis summary for all type consensus truth using LIDC–IDRI(-) data.

Consensus criterion	Data sets (number of cases)	TR		TRE	Automated-Hybrid			
		Overlap (%)	Sphere cases	Overlap (%)	Overlap (%)	TR cases	Sphere cases	TRE cases
25%	Training (300)	66.39 ± 13.03	64/300	75.15 ± 6.91	72.33 ± 9.05	156/300	27/156	144/300
	Validation (66)	66.80 ± 10.86	21/66	73.93 ± 8.72	70.89 ± 9.50	33/66	4/33	33/66
	Testing (66)	69.90 ± 15.71	17/66	77.26 ± 6.78	75.83 ± 7.30	35/66	5/35	31/66
50%	Training (300)	72.52 ± 13.13	68/300	80.84 ± 6.36	78.36 ± 7.71	170/300	33/170	130/300
	Validation (66)	72.59 ± 9.33	21/66	80.07 ± 6.61	76.71 ± 7.75	36/66	8/36	30/66
	Testing (66)	73.85 ± 11.32	25/66	81.06 ± 7.30	77.84 ± 8.25	36/66	9/36	30/66
75%	Training (300)	73.76 ± 14.06	73/300	82.12 ± 6.87	78.52 ± 9.18	208/300	42/208	92/300
	Validation (66)	74.41 ± 14.62	17/66	81.54 ± 7.60	78.75 ± 9.47	45/66	3/45	21/66
	Testing (66)	71.70 ± 19.89	25/66	81.60 ± 8.33	77.93 ± 10.05	45/66	12/45	21/66
100%	Training (300)	70.19 ± 16.15	48/300	79.35 ± 7.73	77.87 ± 8.39	161/300	8/161	139/300
	Validation (66)	69.19 ± 14.41	13/66	79.21 ± 7.97	76.33 ± 9.07	34/66	1/34	32/66
	Testing (66)	68.44 ± 19.71	21/66	80.12 ± 8.07	78.54 ± 8.70	34/66	6/34	32/66

system, operating on 50% consensus truth, yields an overlap of 73.85% on the testing data. The limiting sphere is deployed in 25/66 cases. Using the $\widetilde{TRE}$, the performance rises to 81.06% on the same testing data. The automated hybrid deploys the $\widetilde{TRE}$ engine in 30/66 cases and uses the TR engine in the remaining cases. The overlap score for this hybrid is 77.84%. It is interesting to note that the overlap performances for all of the segmentation systems is very similar across the training, validation, and testing subsets. This gives us confidence that the systems are not overtrained.

We also analyzed how well the segmentation engines perform if the parameter selection was ideal. This lets us evaluate how much performance loss is due to the limitations of the segmentation engines, and how much can be attributed to the imperfect RNN parameter selection process. The ideal/optimum parameter selection for the segmentation engines can be done by conducting an exhaustive search over a fine grid of parameters while making use of (relative to) the provided consensus truth (i.e., a "truth driven" search as opposed to "RNN driven" search). This analysis shows that using the optimum TR parameters for each nodule in the testing set would yield an overlap score of 81.43% for 50% consensus truth. Thus, the TR engine is capable of exceptional performance. The loss of performance due to the imperfect RNN parameter selection is 7.58%. In the case of $\widetilde{TRE}$, the engine with ideal parameters would produce an overlap score on the testing data of 86.30% for 50% consensus truth. The loss due to imperfect

RNN parameter selection is 5.24%. While the RNNs do not provide perfect parameter select, we believe that their performance is very good. We attribute this largely to the rich set of salient features used here.

5.2. Performance on LIDC

To compare the performance of the proposed systems to previously published methods, we test on the original LIDC dataset, as described in Section 2.2. This set is comprised of 77 nodules that have been segmented by 3 or more radiologists. Since previous methods used only the 50% consensus truth, these are the results we present here for our systems as well. The TR and $\widetilde{TRE}$ systems are trained on LIDC–IDRI(-) using the training and validation sets with 50% consensus truth. The results are summarized in Table 7. Our proposed methods are compared to several other published results using the same LIDC data. This comprehensive comparison is made possible thanks to the results published in Wang et al. (2007, 2009), Kubota et al. (2011). In addition to results for their own algorithms, they report results for four other algorithms (Zhao et al., 1999b; Okada and Akdemir, 2005; Okada et al., 2005; Kuhnigk et al., 2006; Li et al., 2008). As shown in Table 7, the proposed TR system provides the highest overlap scores on LIDC data to date at 69.23%. Like the comparison methods, this method uses only a single centralized cue point. Results for our

Table 7

Nodule segmentation performance comparison on LIDC nodules segmented by three or more radiologists. Overlap scores are relative to the 50% consensus truth.

Systems	Performance (average overlap)
Zhao et al. (1999b), Wang et al. (2009)**	43%
Okada and Akdemir (2005), Okada et al. (2005), Kubota et al. (2011)	45 ± 21%
Kuhnigk et al. (2006), Kubota et al., 2011	56 ± 18%
Wang et al. (2007)**	64%
Tuinstra (2008)*	67 ± 16%
Li et al. (2008), Wang et al. (2009)**	45%
Wang et al. (2009)**	58%
Messay et al. (2010)	63 ± 16%
Kubota et al. (2011)	59 ± 19%
Proposed Fully Automated Approach (2014)	
TR50	69.23 ± 13.82%
Proposed Semi-Automated Approaches (2014)	
TRE50	77.58 ± 8.63%
Automated-Hybrid50	74.14 ± 9.88%

* The system assumes a lung mask has been supplied and has limitation on the dimensions of the cropped VOI. Hence, it was only tested on 69 nodules (out of 77).
** For those systems a distribution of the overlap scores can be found in the cited documents.

$\widetilde{TRE}$ and hybrid systems are also listed in Table 7. However, these systems should be considered to be part of a new class of algorithms, as they make use of additional control points. It is clear that the SA systems provide a considerable boost in performance here, on par with what we see for LIDC–IDRI(-). This demonstrates what is possible when an expert user provides more than just a single cue point. We believe this is an important area to explore as a means of balancing performance with user workflow considerations for challenging segmentation cases.

Fig. 9 presents a number of output segmentations for individual LIDC nodules, along with the overlap scores. The segmentations are shown at the maximum area slice. As expected, $\widetilde{TRE}$ outperforms TR in most cases. The main limitations of the TR system are inadequate lung masks for some invasive juxtapleural nodules, and "missing slices" due to partial volume effects and noise. In some cases, there is a great intensity discrepancy between the extreme slices (i.e, the "end caps") and the inner slices due to partial volume effects. Those extreme slices are sometimes missed with our TR system. Those two problems are effectively resolved by the $\widetilde{TRE}$ approach, since it makes use of the revised lung mask and the bounding slices are supplied by the expert user. In a few cases, such as the three nodules shown in Fig. 9, the TR system can actually outperform $\widetilde{TRE}$. Other notable results are the two cases shown in Fig. 9, which show our capability of delineating exclusion regions.

5.3. Discussion

We have computed the overlap score among the various types of consensus truth in LIDC–IDRI(-) and we report a confusion of these scores in Table 8. This table illustrates the levels of dissent among the multiple types of consensus truth. It reads as follows. Each value in the table represents the overlap score (reported in percentage) between the two consensus truths indicated by the corresponding row and column headings. Note that the average overlap between the two extremes of consensus truth shown in the table is only 49.55%. In addition to the confusion matrix, we have computed additional related statistics. For example, the minimum and maximum overlap pairs of segmentations among the four board-certified radiologists, averaged over nodules, is $58.45 \pm 12.16\%$ and $79.57 \pm 8.47\%$, respectively. The average overlap between any two different radiologists is $68.23 \pm 8.86\%$. We draw two important conclusions from all of these findings. First, by analyzing the relationship between these different types of truth, we are able to better understand the levels of agreement and disagreement among board certified radiologists. Second, the overlap of some of the different consensus truths, and individual radiologist segmentations, is similar to that of segmentations from top performing automated systems relative to a defined truth.

In a final experiment with the LIDC–IDRI(-) data, we have attempted to compare overlap scores of the TR and $\widetilde{TRE}$ segmentation systems to that of a human radiologist. Note that the truth provided with LIDC–IDRI(-) does not allow us to track a single radiologist across different nodules. In addition, a consensus truth is generally considered to be superior to that of using a single
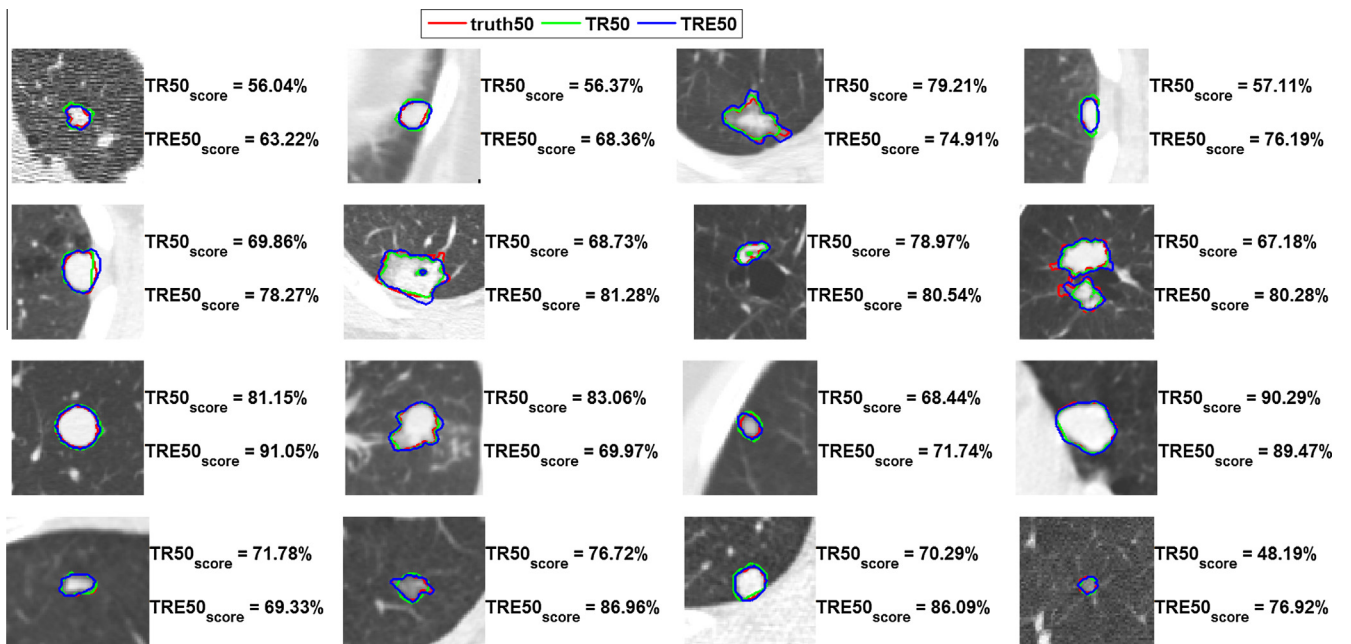

Fig. 9. Segmentation examples for the LIDC nodules generated using the proposed methods.

Table 8
Confusion table (generated using LIDC–IDRI(-)) illustrating levels of dissent among the multiple types of consensus truth.

Consensus	25%	50%	75%	100%
25%	$100.00 \pm 0.00\%$	$77.16 \pm 8.59\%$	$63.73 \pm 10.47\%$	$49.55 \pm 12.01\%$
50%	$77.16 \pm 8.59\%$	$100.00 \pm 0.00\%$	$82.24 \pm 7.65\%$	$63.75 \pm 11.89\%$
75%	$63.73 \pm 10.47\%$	$82.24 \pm 7.65\%$	$100.00 \pm 0.00\%$	$77.13 \pm 10.18\%$
100%	$49.55 \pm 12.01\%$	$63.75 \pm 11.89\%$	$77.13 \pm 10.18\%$	$100.00 \pm 0.00\%$

radiologist as the "gold" standard truth for a performance analysis. Thus, to create the comparison we desire, we have generated a 2 out of 3 radiologist consensus truth for each nodule in the testing set and we score the remaining radiologist against that truth. We repeat this process, scoring each of the four radiologist's segmentations in this way. By averaging these overlap scores we get a performance benchmark for the "average" radiologist, against a 2/3 consensus truth, of $73.49 \pm 11.13\%$. Scoring with the same testing data against the same 2/3 consensus truth, the TR and $\widetilde{T}RE$ segmentation systems trained with 50% consensus truth, provide overlap scores of $72.01 \pm 14.60\%$ and $78.85 \pm 8.30\%$, respectively. Thus, one may conclude that the TR segmentation system performance is close to, but somewhat below, that of the "average" human radiologist. However, the $\widetilde{T}RE$ system outperforms a single independent human radiologist in this metric. We attribute this to the fact that $\widetilde{T}RE$ engine is guided by the 8 control points that are specifically tailored to the desired truth.

One way to exploit the ability of the proposed systems to adapt to different truth is train an array of systems with different truth, as we have done here. Then for a given nodule, a user could be shown all of the system outputs, and could select from among these. Such a system could learn from the preferences of the user, and progressively narrow in on the favored system. Furthermore, it may be possible to retrain the RNN system after a suitable number of nodule segmentations have been "approved" by the user. Thus, radiologist approved segmentation outputs could be used to constitute a new training set to better capture their preferences. Note that we can simulate a hybrid system that is guided by an expert user, as opposed to our automated hybrid system. In this case, we select the best of the TR and $\widetilde{T}RE$ outputs based on truth, rather than using the automated decision rule described in Section 3.3. In this case we are able to achieve a $79.01 \pm 8.56\%$ on LIDC, which is a boost of 4.87% over the automated hybrid, and a boost of 1.43% over using the $\widetilde{T}RE$ for all nodules. This suggests that there could be benefits to presenting the user with multiple outputs to select from.

A valuable area of future work would be to focus on implementation, and investigate software and hardware acceleration methods to make this method suitable for clinical practice. As mentioned in Section 4.1, our focus here is not on processing speed, but rather segmentation performance. However, the reader may be interested to know what the processing speed is for the current prototype MATLAB implementation using no hardware acceleration of nay kind. Generating 8 outputs, for all 8 systems listed in Table 3, takes approximately 4 min per nodule on an HP Personal Computer (PC) with processor speed of 3.22 GHz. This time excludes the automated lung segmentation algorithm described in Messay et al. (2010) that consumes approximately 27–35 s per CT scan. The main computational burden comes in calculating the features for each candidate segmentation when searching the discrete segmentation engine parameter space. To reduce the number of candidates to evaluate, efficient search strategies may be employed (Tuinstra, 2008). However, one can expect a slight reduction in performance using fast searches as the globally optimum solution may not be found. Even with a simple grid search, processing can be accelerated using parallel processing (Messay et al., 2011). This is because the features for each segmentation candidate can be computed independently of one another. Another potentially fruitful area of future research might be a study on the sensitivity of the subsystems to the supplied cue points. These are matters we intend to examine in future work.

6. Conclusions

In this paper, we have presented new pulmonary nodule segmentation algorithms. These include the FA system that uses our TR segmentation engine, the SA system employing the $\widetilde{T}RE$ segmentation engine, and a hybrid system that uses both. The TR segmentation engine is a traditional automated system that requires only a single user supplied cue point. The $\widetilde{T}RE$ engine is semi-automated in the sense that 8 user supplied control points are required. While the burden on the user is certainly greater with the $\widetilde{T}RE$ engine, the resulting system is highly robust and can handle a variety of challenging cases. The hybrid system attempts to balance the ease of use of the TR segmentation engine with the performance boost of the $\widetilde{T}RE$ engine. This hybrid system initially only requires a single cue point. Only if the resulting segmentation is determined to be inadequate is the user prompted to enter the 8 control points for the $\widetilde{T}RE$ engine.

To the best of our knowledge, the results summarized in Table 6 represent one of the first performance benchmarks using the new LIDC-IDLI dataset. Note that both testing data subsets used here to evaluate our FA, SA and hybrid systems are publicly available via the links provided in Section 2. It is our hope that this benchmark will spark future research efforts. We also compare the performance of the proposed methods with several previously reported results using the same LIDC data and performance metric. These results, summarized in Table 7, show that the TR system provides the highest overlap scores of those systems considered. In particular, the TR system generates overlap scores that are 5.34% above the next best results (excluding results from one of the current authors). The results also show that the $\widetilde{T}RE$ engine provides a boost of 8.35% over the TR engine on LIDC, and a similar boost on LIDC–IDRI(-). Thus, the 8 user provided cue points are clearly an effective way to boost nodule segmentation system performance. We believe that such semi-automated systems represent a new class of nodule segmentation algorithms capable of performing at new levels, and they should continue to be explored.

References

ACS, 2013. American Cancer Society, Cancer Facts and Figures. <http://www.cancer.org/research/cancerfactsstatistics/index>.

Armato III, S., Giger, M., MacMahon, H., 2001. Automated detection of lung nodules in CT scans: preliminary results. Med. Phys. 28, 1552–1561.

Armato III, S., McLennan, G., McNitt-Gray, M., Meyer, C., Yankelevitz, D., Aberle, D., Henschke, C., Hoffman, E., Kazerooni, E., MacMahon, H., et al., 2004. Lung Image Database Consortium: developing a resource for the medical imaging research community. Radiology 232, 739.

Armato III, S., McNitt-Gray, M., Reeves, A., Meyer, C., McLennan, G., Aberle, D., Kazerooni, E., MacMahon, H., van Beek, E., Yankelevitz, D., et al., 2007. The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. Acad. Radiol. 14, 1409–1421.

Armato III, S.G., Giger, M.L., Moran, C.J., Blackburn, J.T., Doi, K., MacMahon, H., 1999. Computerized detection of pulmonary nodules on CT scans1. Radiographics 19, 1303–1311.

Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al., 2011. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Med. Phys. 38, 915–931.

Aubin, C.É., Descrimes, J., Dansereau, J., Skalli, W., Lavaste, F., Labelle, H., et al., 1994. Geometrical modeling of the spine and the thorax for the biomechanical analysis of scoliotic deformities using the finite element method. In: Annales de chirurgie, pp. 749–761.

Bendtsen, C., Kietzmann, M., Korn, R., Mozley, P.D., Schmidt, G., Binnig, G., 2011. X-ray computed tomography: semiautomated volumetric analysis of late-stage lung tumors as a basis for response assessments. J. Biomed. Imaging 2011, 5.

Coleman, T.F., Li, Y., Mariano, A., 1998. Segmentation of Pulmonary Nodule Images Using Total Variation Minimization. Technical Report. Cornell University.

Dehmeshki, J., Amin, D., Valdivieso, M., Ye, X., 2008. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. IEEE Trans. Med. Imaging 27, 467–480. http://dx.doi.org/10.1109/TMI.2007.907555.

Diciotti, S., Picozzi, G., Falchini, M., Mascalchi, M., Villari, N., Valli, G., 2008. 3-d segmentation algorithm of small lung nodules in spiral CT images. Trans. Info. Tech. Biomed. 12, 7–19. http://dx.doi.org/10.1109/TITB.2007.899504, <http://dx.doi.org/10.1109/TITB.2007.899504>.

Diepenbrock, S., Ropinski, T., 2012. From imprecise user input to precise vessel segmentations. In: Eurographics Workshop on Visual Computing for Biology and Medicine, The Eurographics Association, pp. 65–72.

Elmoataz, A., Schuepp, S., Bloyet, D., 2001. Fast and simple discrete approach for active contours for biomedical applications. Int. J. Pattern Recognit. Artif. Intell. 15, 1201–1212.

Ernst, R., Hardie, R., Gurcan, M., Oto, A., Rogers, S., Hoffmeister, J., 2004. CAD Performance Analysis for Pulmonary Nodule Detection: Comparison of Thick- and Thin-Slice Helcial CT. Radiology Society of North America (RSNA).

Fan, L., Qian, J., Odry, B., Shen, H., Naidich, D., Kohl, G., Klotz, E., 2002. Automatic segmentation of pulmonary nodules by using dynamic 3d cross-correlation for interactive CAD systems. In: Proc. SPIE, pp. 1362–1369.

Giger, M.L., Ahn, N., Doi, K., MacMahon, H., Metz, C.E., 1990. Computerized detection of pulmonary nodules in digital chest images: use of morphological filters in reducing false-positive detections. Med. Phys. 17, 861–865.

Giger, M.L., Bae, K.T., MacMahon, H., 1994. Computerized detection of pulmonary nodules in computed tomography images. Invest. Radiol. 29, 459–465.

Giger, M.L., Doi, K., MacMahon, H., 1988. Image feature analysis and computer-aided diagnoses in digital radiography: automated detection of nodules in peripheral lung fields. Med. Phys. 15, 158–166.

van Ginneken, B., 2001. Computer-Aided Diagnosis in Chest Radiography. Ph.D. Thesis. Utrecht University, The Netherlands.

van Ginneken, B., 2006. Supervised probabilistic segmentation of pulmonary nodules in CT scans. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006. Springer, pp. 912–919.

van Ginneken, B., Frangi, A.F., Staal, J.J., ter Haar Romeny, B.M., Viergever, M.A., 2002. Active shape model segmentation with optimal features. IEEE Trans. Med. Imaging 21, 924–933.

van Ginneken, B., Stegmann, M., Loog, M., 2006. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med. Image Anal. 10, 19–40.

Girosi, F., Jones, M., Poggio, T., 1995. Regularization theory and neural networks architectures. Neural Comput. 7, 219–269.

Gu, Y., Kumar, V., Hall, L.O., Goldgof, D.B., Li, C.Y., Korn, R., Bendtsen, C., Velazquez, E.R., Dekker, A., Aerts, H., et al., 2013. Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. Pattern Recognit. 46, 692–702.

Gurcan, M., Hardie, R., Rogers, S., 2004. Shape Estimates and Temporal Registration of Lesions and Nodules. US Patent App. 10/993,176.

Ham, F.M., Kostanic, I., 2000. Principles of Neurocomputing for Science and Engineering. McGraw-Hill Higher Education.

Hardie, R., Rogers, S., Wilson, T., Rogers, A., 2008. Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. Med. Image Anal. 12, 240–258. http://dx.doi.org/10.1016/j.media.2007.10.004.

Henschke, C.I., Yankelevitz, D.F., Mirtcheva, R., McGuinness, G., McCauley, D., Miettinen, O.S., 2002. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. Am. J. Roentgenol. 178, 1053–1057.

Horsthemke, W.H., Raicu, D.S., Furst, J.D., 2010. Predicting LIDC diagnostic characteristics by combining spatial and diagnostic opinions. In: SPIE Medical Imaging, International Society for Optics and Photonics, pp. 76242Y–76242Y.

Jacobs, C., van Rikxoort, E.M., Twellmann, T., Scholten, E.T., de Jong, P.A., Kuhnigk, J.M., Oudkerk, M., de Koning, H.J., Prokop, M., Schaefer-Prokop, C., et al., 2014. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. Med. Image Anal. 18, 374–384.

Kawata, Y., Niki, N., Ohamatsu, H., Kusumoto, M., Kakinuma, R., Mori, K., Nishiyama, H., Eguchi, K., Kaneko, M., Moriyama, N., 2003. Pulmonary nodule segmentation in thoracic 3d CT images integrating boundary and region information. In: Medical Imaging 2003, International Society for Optics and Photonics, pp. 1520–1530.

Keshani, M., Azimifar, Z., Tajeripour, F., Boostani, R., 2013. Lung nodule segmentation and recognition using svm classifier and active contour modeling: a complete intelligent system. Comput. Biol. Med. 43, 287–300.

Ko, J.P., Rusinek, H., Jacobs, E.L., Babb, J.S., Betke, M., McGuinness, G., Naidich, D.P., 2003. Small pulmonary nodules: volume measurement at chest CT-phantom study. Radiology 228, 864–870. http://dx.doi.org/10.1148/radiol.2283020059.

Korfiatis, P., Skiadopoulos, S., Sakellaropoulos, P., Kalogeropoulou, C., Costaridou, L., 2014. Combining 2D Wavelet Edge Highlighting and 3D Thresholding for Lung Segmentation in Thin-Slice CT.

Kosko, B., 1992. Neural Networks and Fuzzy Systems. Prentice Hall.

Kostis, W.J., Reeves, A.P., Yankelevitz, D.F., Henschke, C.I., 2003. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Trans. Med. Imaging 22, 1259–1274.

Kubota, T., Jerebko, A., Salganicoff, M., Dewan, M., Krishnan, A., 2008. Robust segmentation of pulmonary nodules of various densities: from ground-glass opacities to solid nodules. In: Proceedings of the International Workshop on Pulmonary Image Processing, pp. 253–262.

Kubota, T., Jerebko, A.K., Dewan, M., Salganicoff, M., Krishnan, A., 2011. Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models. Med. Image Anal. 15, 133–154.

Kuhnigk, J.M., Dicken, V., Bornemann, L., Bakai, A., Wormanns, D., Krass, S., Peitgen, H.O., 2006. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. IEEE Trans. Med. Imaging 25, 417–434.

Lay, D., 2003. Linear Algebra and its Applications. Pearson Education, Inc.

Li, Q., Li, F., Doi, K., 2008. Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier. Acad. Radiol. 15, 165–175.

McNitt-Gray, M.F., Armato III S.G., Meyer, C.R., Reeves, A.P., McLennan, G., Pais, R.C., Freymann, J., Brown, M.S., Engelmann, R.M., Bland, P.H., et al., 2007. The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. Acad. Radiol. 14, 1464–1474.

Messay, T., Chen, C., Ordóñez, R., Taha, T.M., 2011. GPGPU acceleration of a novel calibration method for industrial robots. In: Proceedings of the 2011 IEEE National Aerospace and Electronics Conference (NAECON). IEEE, pp. 124–129.

Messay, T., Hardie, R.C., Rogers, S.K., 2010. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med. Image Anal. 14, 390. http://dx.doi.org/10.1016/j.media.2010.02.004.

Mitton, D., Landry, C., Veron, S., Skalli, W., Lavaste, F., De Guise, J., 2000. 3d reconstruction method from biplanar radiography using non-stereocorresponding points and elastic deformable meshes. Med. Biol. Eng. Comput. 38, 133–139.

Moltz, J.H., Bornemann, L., Kuhnigk, J.M., Dicken, V., Peitgen, E., Meier, S., Bolte, H., Fabel, M., Bauknecht, H.C., Hittinger, M., Kießling, A., Pusken, M., Peitgen, H.O., 2009. Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in CT scans. IEEE J. Sel. Topics Signal Process. 3, 122–134. http://dx.doi.org/10.1109/jstsp.2008.2011107, <http://dx.doi.org/10.1109/jstsp.2008.2011107>.

Moshtagh, N., 2005. Minimum volume enclosing ellipsoid. Convex Optim.

Moura, D.C., Boisvert, J., Barbosa, J.G., Tavares, J.M.R., 2009. Fast 3d reconstruction of the spine using user-defined splines and a statistical articulated model. In: Advances in Visual Computing. Springer, pp. 586–595.

Mullally, W., Betke, M., Wang, J., Ko, J.P., 2004. Segmentation of nodules on chest computed tomography for growth assessment. Med. Phys. 31, 839.

Okada, K., Akdemir, U., 2005. Blob segmentation using joint space-intensity likelihood ratio test: application to 3d tumor segmentation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE, pp. 437–444.

Okada, K., Comaniciu, D., Krishnan, A., 2005. Robust anisotropic gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT. IEEE Trans. Med. Imaging 24, 409–423.

Opfer, R., Wiemker, R., 2007. Performance analysis for computer-aided lung nodule detection on LIDC data. In: Jiang Yulei, Sahiner, Berkman (Eds.), Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment. Proceedings of the SPIE, 6515, 65151C.

Papoulis, A., Pillai, S.U., 2002. Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York.

Pham, D.L., Xu, C., Prince, J.L., 2000. Current methods in medical image segmentation 1. Annu. Rev. Biomed. Eng. 2, 315–337.

Pomero, V., Mitton, D., Laporte, S., de Guise, J.A., Skalli, W., 2004. Fast accurate stereoradiographic 3d-reconstruction of the spine using a combined geometric and statistic model. Clin. Biomech. 19, 240–247.

Reeves, A., Biancardi, A., Apanasovich, T., Meyer, C., MacMahon, H., van Beek, E., Kazerooni, E., Yankelevitz, D., McNitt-Gray, M., McLennan, G., et al., 2007. The Lung Image Database Consortium (LIDC) a comparison of different size metrics for pulmonary nodule measurements. Acad. Radiol. 14, 1475–1485.

Reeves, A., Jirapatnakul, A., Biancardi, A., Apanasovich, T., Schaefer, C., Bowden, J., Kietzmann, M., Korn, R., Dillmann, M., Li, Q., et al., 2009. The Volcano09 Challenge: Preliminary Results. VOLCANO09, pp. 353–364.

Reeves, A.P., Chan, A.B., Yankelevitz, D.F., Henschke, C.I., Kressler, B., Kostis, W.J., 2006. On measuring the change in size of pulmonary nodules. IEEE Trans. Med. Imaging 25, 435–450.

Rogers, S.K., 1991. An Introduction to Biological and Artificial Neural Networks for Pattern Recognition, vol. 4. SPIE Press.

Rousson, M., Bai, Y., Xu, C., Sauer, F., 2006. Probabilistic minimal path for automated esophagus segmentation, in: Medical Imaging, International Society for Optics and Photonics, pp. 614449–614449.

Sahiner, B., Hadjiiski, L., Chan, H., Shi, J., Cascade, P., Kazerooni, E., Zhou, C., Wei, J., Chughtai, A., Poopat, C., et al., 2007. Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: observer performance study. In: Proceedings of SPIE 6515, 65151D.

Serra, J., 1983. Image Analysis and Mathematical Morphology. Academic Press, Inc., Orlando, FL, USA.

Strickland, R., 2002. Image-Processing Techniques for Tumor Detection. CRC Press.

Tachibana, R., Kido, S., 2006. Automatic segmentation of pulmonary nodules on CT images by use of nci lung image database consortium. In: Medical Imaging, International Society for Optics and Photonics, pp. 61440M–61440M.

Tuinstra, T.R., 2008. Automatic Segmentation of Small Pulmonary Nodules in Computed Tomography Data Using a Radial Basis Function Neural Network with Application to Volume Estimation. Ph.D. Thesis. University of Dayton.

Verikas, A., Bacauskiene, M., 2002. Feature selection with neural networks. Pattern Recogn. Lett. 23, 1323–1335.

Vidal, C., Beggs, D., Younes, L., Jain, S.K., Jedynak, B., 2011. Incorporating user input in template-based segmentation. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, pp. 1434–1437.

Wang, J., Engelmann, R., Li, Q., 2007. Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique. Med. Phys. 34, 4678–4689, <http://view.ncbi.nlm.nih.gov/pubmed/18196795>.

Wang, Q., Song, E., Jin, R., Han, P., Wang, X., Zhou, Y., Zeng, J., 2009. Segmentation of lung nodules in computed tomography images using dynamic programming and multidirection fusion techniques. Acad. Radiol. 16, 678–688.

Wang, W., Jones, P., Partridge, D., 2000. Assessing the impact of input features in a feedforward neural network. Neural Comput. Appl. 9, 101–112.

Way, T.W., Hadjiiski, L.M., Sahiner, B., Chan, H.P., Cascade, P.N., Kazerooni, E.A., Bogot, N., Zhou, C., 2006. Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3d active contours. Med. Phys. 33, 2323.

Wiemker, R., Zwartkruis, A., 2001. Optimal thresholding for 3d segmentation of pulmonary nodules in high resolution CT. In: Lemke, H.U., Vannier, M.W., Inamura, K., Farman, A.G., Doi, K. (Eds.), CARS 2001. Computer Assisted Radiology and Surgery. Proceedings of the 15th International Congress and Exhibition, June 27–30, 2001. Elsevier, Berlin, Germany, pp. 653–658.

Wormanns, D., Diederich, S., 2004. Characterization of small pulmonary nodules by CT. Eur. Radiol. 14, 1380–1391.

Xu, N., Ahuja, N., Bansal, R., 2002. Automated lung nodule segmentation using dynamic programming and em-based classification. In: Medical Imaging 2002, International Society for Optics and Photonics, pp. 666–676.

Ye, X., Siddique, M., Douiri, A., Beddoe, G., Slabaugh, G., 2009. Image segmentation using joint spatial-intensity-shape features: application to CT lung nodule segmentation. In: SPIE Medical Imaging, International Society for Optics and Photonics, pp. 72594V–72594V.

Zhao, B., Reeves, A.P., Yankelevitz, D.F., Henschke, C.I., 1999a. Three-dimensional multicriterion automatic segmentation of pulmonary nodules of helical computed tomography images. Opt. Eng. 38, 1340–1347. http://dx.doi.org/10.1117/1.602176, +http://dx.doi.org/10.1117/1.602176.

Zhao, B., Yankelevitz, D., Reeves, A., Henschke, C., 1999b. Two-dimensional multi-criterion segmentation of pulmonary nodules on helical CT images. Med. Phys. 26, 889.