Lung nodule segmentation via level set machine learning

Matthew C Hancock, Jerry F Magnan
Florida State University, Department of Mathematics, 208 Love Building, 1017 Academic Way, Tallahassee, FL 32306, USA

Abstract. Lung cancer has the highest mortality rate of all cancers in both men and women. The algorithmic detection, characterization, and diagnosis of abnormalities found in chest CT scan images can potentially aid radiologists by providing additional medical information to consider in their assessment. Lung nodule segmentation, i.e., the algorithmic delineation of the lung nodule surface, is a fundamental component of an automated nodule analysis pipeline. We introduce an extension of the vanilla level set image segmentation method where the velocity function is learned from data via machine learning regression methods, rather than manually designed. This mitigates the tedious design process of the velocity term from the standard method. We apply the method to image volumes of lung nodules from CT scans in the publicly available LIDC dataset, obtaining an average intersection over union score of 0.7185 ($\pm$0.1114).

Keywords: lung nodule segmentation, level set method, machine learning, computer-aided diagnosis.

1 Introduction

Lung cancer has the highest mortality rate in both males and females in the United States. A lung nodule is a small to medium sized (roughly, 3 mm to 30 mm) abnormal region with a somewhat well-defined boundary. The definition is inherently imprecise because it is based on visual examples and the subjective interpretations thereof. The likelihood of malignancy of a lung nodule can be inferred by a combination of radiological features (e.g., growth rate, shape, or density features), which if determined early, increases chances of survival. The precise location of the nodule’s surface in the image volume is often necessary to produce such nodule features computationally (e.g., a nodule’s volume or average internal density). Thus, the accurate segmentation of the nodule is a crucial step in a computer-aided diagnosis (CAD) system. However, lung nodule segmentation is challenging because of the variability in nodule appearance and shape, as well as the potential proximity to various other lung anatomy (e.g., the vasculature or the pleural wall).

In this work, we introduce a machine learning extension of the standard level set image segmentation method of Malladi and Sethian and apply it to the lung nodule segmentation problem. Starting from an initial guess, the method evolves a function that moves toward the boundary of the lung nodule in the image volume. The evolution of this function is dictated by a partial differential equation (PDE) whose velocity term is a function of the underlying image data, thus guiding the surface towards the desired boundary. The standard level set approach requires the manual design of this velocity term, which is difficult. In our extension of this method, which we call the level set machine learning (LSML) method, the velocity term in the PDE is learned from data via machine learning regression models.

This paper is structured as follows: in Section 2 we provide background on the level set segmentation method, related works and results for lung nodule segmentation. In Section 3, we introduce the LSML image segmentation method, and in Section 4, we present and discuss our results from applying our method to the lung nodule segmentation problem. We conclude in Section 5.
Fig 1: The surface is an evolving function \( u \) whose zero level curve undergoes a topological change between \( t = 3/4 \) and \( t = 1 \).

2 Background

2.1 Level set image segmentation and variants

Segmentation of images by evolving contours was introduced by Kass et al.\(^6\) where a parameterized curve is evolved by minimizing a weighted sum of internal and external energy functionals. Evolving a parametrized curve is tedious from a computational point of view because topological changes are not easily handled. On the other hand, implicit curves handle topological changes very naturally. In the level set method, the curve denoting an object’s boundary is given implicitly by the zero level curve of a function \( u \) which is often called the level set function. Figure 1 illustrates this concept, where evolving the surface \( u \) yields a zero level curve (shown in red below the surface \( u \)) that undergoes a change in number of connected components from one to two. The concept extends readily to higher dimensions where the zero level set is a surface. The level set approach was pioneered by Osher and Sethian\(^7\) for tracking flame movement, and Malladi and Sethian\(^5\) introduced the level set method to the realm of image segmentation.

The movement of the level set function \( u \) is governed by a PDE that we briefly derive. Suppose \( x(q) \) parameterizes the zero level set, \( \{ u = 0 \} \), and evolves in the outward normal direction with velocity \( \nu(x) \), i.e., \( x_t = \nu N \), where \( x_t \) is the time partial derivative of \( x(q,t) \) and \( N \) is the outward unit normal to the zero level set. Thus, positive values of \( \nu \) expand the level set and negative values of \( \nu \) contract it. We take the convention that \( u \) is positive inside the zero level set and negative outside, so that the gradient vector \( Du = (u_{x_1}, \ldots, u_{x_n}) \) points in the inward normal direction along the zero level set of \( u \) and \( N = -Du / \| Du \| \). Thus by differentiating the level set relation \( u(x(t), t) = 0 \) with respect to \( t \), we arrive at the level set evolution PDE,

\[
\frac{\partial u}{\partial t} = \mu \| Du \|
\] (1)

To perform image segmentation, the velocity \( \nu \) in Eq. (1) is defined in terms of the underlying image information. Intuitively, the velocity should be positive interior to the target boundary to cause expansion and negative exterior to the boundary to cause contraction. Near the target boundary, the velocity should be small in magnitude to prevent overshooting. The manual design of such a velocity field is difficult and often entails simplifying assumptions such as a homogeneously bright object against a homogeneously dark background. Such assumptions are often violated in practice: boundaries can be fuzzy, occluded, or defined in terms other than those assumed. These difficulties often occur in medical imagery where lung nodules are attached, or in close proximity, to separate anatomical objects with locally similar appearance, e.g., juxta-pleural or juxta-vascular lung nodules. The result of faulty assumptions is that the zero level set fails to evolve towards the desired
In typical level set image segmentation approaches, statistical information garnered from a dataset of training examples is not used. A few works have explored this avenue, mostly by attempting to enforce the evolving function \( u \) to conform to more statistically likely shapes or to enclose more image regions with more statistically like appearance. The first work to introduce such an approach is due to Leventon, where principal component analysis (PCA) was used on a dataset of training images and shapes to create Bayesian prior models in order to penalize segmentations that deviate from those with expected shape and expected image values near the segmentation surfaces. Tsai took a similar approach by applying PCA to a training set of signed distance representation of shapes. More recent work has formulated the level set evolution as an energy functional minimization problem, where statistical information about shape and image features is incorporated by viewing the energy functional as the negative log of some probability density. The probability density of image and shape features is modeled by employing, for example, Gaussian kernel density estimation.

The work closest in spirit to our approach is not level set based, but rather Van Ginneken’s machine learning extension of the standard region growing method. Traditionally, the region growing method recursively adds points to a growing region via a fixed rule (e.g., based on an image value threshold); however, Van Ginneken allows this rule to be learned from data, viewing the choice of whether or not to add a point as a binary classification problem solvable via machine learning methods.

### 2.2 Lung Nodule Image Segmentation

Lung nodule image segmentation refers to the algorithmic delineation of the boundaries of objects called lung nodules, which are focal abnormalities of the lung, often appearing as dense regions relative to their surroundings. Lung nodule image segmentation is difficult because of the variation in nodule geometry, variation in nodule interior- and edge-densities, and non-nodule anatomical structures in close proximity to, or occluding, a nodule, which often have similar density to the nodule. Most lung nodule segmentation work, including our own, leverages the publicly available LIDC dataset of lung CT data and radiologist annotations. The LIDC dataset contains 1018 lung CT scans that have been annotated by four radiologists (see Fig. 2 for an example). Each radiologist visually examined each scan, and upon detection of a lung nodule within a scan, drew the boundary of the lung nodule in each slice for which the detected nodule was present (according to that specific radiologist). These “ground-truth” nodule boundary annotations, along with CT image volume data, are available in the LIDC dataset.

An assortment of methods have been applied to the lung nodule segmentation problem. The recent work by Wang constructed a table of works, including their own, that reported the Jaccard overlap score. The Jaccard score is also referred to as the “intersection over union” score and is a measure of segmentation quality. It is defined as the size of the intersection between the algorithmic segmentation and the ground-truth segmentation divided by the size of the union, which is one for perfect overlap and zero at worst when there is no overlap. Because we have also used the Jaccard overlap score as the measure of segmentation quality in our work, we have included and expanded this list of works in Table 1, where we have placed our work in context for comparison with these other methods: Tachibana combined a variety of image processing techniques such as
thresholding, template-matching, and the watershed method (an edge-based method for determining boundaries) and obtained an average Jaccard overlap score of 0.5070 on 23 nodules. Wang\textsuperscript{16} used a dynamic programming approach and fusion method for combining information from multiple two-dimensional image slices. They reported an average Jaccard overlap score of 0.58 on 64 nodules. Messay\textsuperscript{17,18} in their first work, applied a variety of morphological operations with a subsequent “rule-based analysis”, obtaining average overlap of 0.63 over 68 nodules, whereas in their follow-up work, they used a calibration process over training data to predict various thresholding and morphological parameters based on features computed from the image, improving their results to an average of 0.7170 over 66 nodules. Both Kubota\textsuperscript{19} and Lassen\textsuperscript{20} applied basic image processing techniques such as thresholding and morphological operations, as well as convexity information, achieving average Jaccard overlap scores of 0.69 and 0.52, respectively. Tan\textsuperscript{21} used the watershed method, active contours, and Markov-random fields, achieving an average overlap score of 0.65 over a dataset containing 23 nodules. The work by Wang\textsuperscript{14} applied convolutional neural networks (CNNs) with a centrally-focused max-pooling operation applied to 493 test nodules, obtaining an average overlap score of 0.7116.

Relatively fewer works have applied the level set method for lung nodule image segmentation. Schildkraut\textsuperscript{22} tested the level set method with energy terms, for “increasing contrast of the segmented region relative to its surroundings”, to 23 lung nodules in radiography images. They reported an average overlap using the Dice coefficient, \( S(A, B) = \frac{2|A \cap B|}{|A| + |B|} \), of 0.6477 on the 23 lung nodules. The Dice score is related to the Jaccard score by \( J = \frac{S}{2-S} \), and thus, the work by Schildkraut\textsuperscript{22} reported an Jaccard overlap score of 0.4790. In the work of Tan\textsuperscript{21}, which we have mentioned previously, a level set formulation of the geometric active contours method was employed as a post-processing step following an initial watershed segmentation, achieving an average Jaccard overlap score of 0.65. Farag\textsuperscript{23} used a level set approach and incorporated an elliptical prior to aid in cases where lung nodules are in proximity to other anatomical objects. They reported a “success rate” (where success is determined by visual inspection of the resulting segmentation) of 94.61% on 334 lung nodules images from the LIDC dataset, but the Jaccard overlap score or similar measures of overlap are not reported. Farhangi\textsuperscript{24} used the region-based Chan-Vese\textsuperscript{25} active contour model to partition nodule and non-nodule regions based on region image-homogeneity using a level set formulation. In addition, a training set of nodule shapes was employed, and at each iteration during the level set evolution, the level set iterate was projected onto the linear span
Table 1: Performance of various lung nodule segmentation methods under Jaccard overlap metric when available.

| Authors          | Year | Number of Nodules | Jaccard overlap     |
|------------------|------|-------------------|---------------------|
|                  |      | Training | Testing |                     |
| Tachibana\textsuperscript{15} | 2006 | -         | 23      | 0.5070 (±0.2190)    |
| Schildkraut\textsuperscript{22} | 2009 | -         | 23      | 0.4790              |
| Wang\textsuperscript{16}        | 2009 | 23        | 64      | 0.5800              |
| Messay\textsuperscript{17}      | 2010 | -         | 68      | 0.6300 (±0.1600)    |
| Kubota\textsuperscript{19}      | 2011 | -         | 23      | 0.6900 (±0.1800)    |
| Tan\textsuperscript{21}         | 2013 | -         | 23      | 0.6500              |
| Farag\textsuperscript{23}       | 2013 | -         | 334     | N/A                 |
| Lassen\textsuperscript{20}      | 2015 | -         | 19      | 0.5200 (±0.0700)    |
| Messay\textsuperscript{18}      | 2015 | 300       | 66      | 0.7170 (±0.1989)    |
| Farhangi\textsuperscript{24}    | 2017 | 488       | 54      | 0.5700 (±0.1600)    |
| Wang\textsuperscript{14} (Level set) | 2017 | 350    | 493     | 0.4350 (±0.0952)    |
| Wang\textsuperscript{14} (CNN)   | 2017 | 350       | 493     | 0.7116 (±0.1222)    |
| LSML method (our work)           | -    | 672       | 112     | 0.7185 (±0.1114)    |

of the training shapes by solving a minimization problem that included a sparsity-inducing term to force coefficients in the weighted sum to be sparse. They used 542 lung nodules from the LIDC dataset, achieving an average Jaccard overlap score of 0.57 over a 10-fold cross-validation procedure. For comparison against their convolutional network model, Wang\textsuperscript{14} also applied a generic version (i.e., non-statistical and without specific tailoring to the lung nodule problem domain) of the region-based Chan-Vese level set model,\textsuperscript{25} obtaining an average overlap score of 0.4350 over the same 493 nodules on which they tested their network model. In our work, we obtain an average Jaccard overlap score of 0.7185 over a testing dataset of 112 nodules using our machine learning extension of the level set method described in Section 3.

3 Methods

We provide motivation for the LSML method in Section 3.1 and an algorithmic outline of the parameter tuning process in Section 3.2. In Section 3.4 we describe our initialization routine that yields a first guess of the segmentation given an image volume containing a lung nodule. In Section 3.5 we describe the features that are used as inputs to the regression models in the LSML method used in our experiments.

3.1 Motivation

Let’s suppose we find ourselves in the not unusual situation of having a dataset of pairs \((M_l, c_l)\), where \(M\) is an image and \(c\) is a curve or surface annotating the boundary of the object to be segmented in \(M\). From the outset, it is not entirely apparent how to incorporate such labeled data to model the velocity term \(\nu\) in Eq. (1); however, we leverage an observation from Breen and Whitaker:\textsuperscript{26} if \(\nu\) is the signed distance transform of the target boundary \(c\), then the zero set of \(u\) converges to \(c\) under the motion dictated by Eq. (1).
This result is intuitive because we’re dictating that the zero level set move towards the \( c \) with speed equal to the distance from the boundary, where the choice of expansion or contraction is controlled by the sign in the signed distance. A bit more formally, this can be understood as follows: first, we define a success score of \( \mathcal{L}(t) = \int H(u) \nu \, dx \), with \( H \) being the unit step function. \( \mathcal{L}(t) \) is maximal when the zero level set of \( u \) matches that of \( \nu \) because otherwise the integral would include a negative part of \( \nu \). Next, we observe the evolution of the score \( \mathcal{L} \) by differentiating with respect to \( t \) and plugging in Eq. (1):

\[
\frac{d\mathcal{L}}{dt} = \int \delta(u) \|Du\| \nu^2 \, dx = \oint_{\{u=0\}} \nu^2 \, ds \geq 0,
\]

which is stationary when the level set of \( u \) matches the zero level set of \( \nu \), thus matching the target curve \( c \).

Our motivation thus far is circular: setting \( \nu \) to the signed distance transform of the annotating curve \( c \) assumes the solution! However, these observations strongly suggest an approximation scheme of the form \( V \approx \nu \), where \( V \) is a machine learning regression model calibrated from image and shape data to approximate the signed distance values from the annotating curves provided by the dataset, thus guiding the level set evolution towards the correct segmentation in cases where the solution is unknown. This is the essence of the LSML approach.

### 3.2 Outline of the LSML method

First, Eq. (1) is discretized:

\[
u_{ijk}^{n+1} = \nu_{ijk}^n + \Delta t \nu_{ijk} \nabla_{n} \]

where the discrete gradient norm term \( \nabla \approx \|Du\| \) is approximated using the upwind scheme of Osher and Sethian.\(^7\) Next, the velocity term is replaced with an iteration-dependent approximation by a regression model \( V^n \approx \nu \), thus replacing Eq. (2) with

\[
u_{ijk}^{n+1} = \nu_{ijk}^n + \Delta t V_{ijk}^n \nabla_{n}
\]

The input to the regression model \( V^n \) is a feature vector computed from image and shape data, thus \( V_{ijk}^n = V_{ijk}^n (F_{ijk}) \), where \( F = F(u, M) \). In other words, the feature vector, which is the input to the regression model \( V^n \), is a function of the image and the level set iterate, from which image and shape features can be extracted, respectively. As a simple example, consider a two-component feature map function that yields two features, \( F_{ijk}(u, M) = \left[ \sum_{qrs} H(u_{qrs}), M_{ijk} \right] \), where \( H \) is the unit step function. The first feature, which approximates the volume enclosed by the zero level set of \( u \), does not depend on the local \((i,j,k)\) position and is thus a global shape feature. The second feature depends on both the local spatial grid index and the image and is therefore a local image feature. Generally, the feature map function may comprise an assortment of combinations of local and global, shape and image features.

The goal of the regression model \( V^n \) is to learn a mapping from the feature vector \( F_{ijk} \), which depends only on local and global image and shape information, to the signed distance value because, as discussed in Section 3.1, these values are known to guide the evolving zero level set towards the target shape. Because there is a regression model for each discrete time-step \( n \), these models are obtained sequentially in the training process. We outline this sequential procedure below, assuming a dataset \((M(l), c(l))_{l=1}^N\) of images and ground-truth boundaries:

1. Initialize each \( u^0(k) \) according to a computationally cheap scheme that yields a first guess of the segmentation for image \( M(k) \). Set \( n = 0 \).
2. Calibrate parameters of regression model \( n \) by least squares over all spatial coordinates and examples, i.e.,

\[
\theta^n = \arg\min_\theta \sum_{ijkl} (V^n_{ijkl}(\cdot | \theta) - v_{ijkl}(l))^2
\]

where \( \theta \) are the parameters of regression model \( V^n \) and \( \nu \) is the signed distance transform of the annotating curve \( c \). Note that we have suppressed the argument to \( V^n \); in full, \( V^n \) is a function of \( F(u^n(l), M(l)) \), i.e., the feature vector for the current level set iterate and image for example \( l \).

3. Step each \( u^n(l) \) forward in time by the PDE discretization Eq. (3)

4. If metrics (e.g., intersection over union) observed over a separate validation set are not degrading, set \( n \leftarrow n + 1 \) and go to Step 2.

The training procedure outlined above yields a sequence of regression models that can be deployed on new, unseen images using the iteration in Eq. (3) with the regression model parameters fixed to those determined from the training procedures outlined above. We note that the procedure can be made much more efficient by considering spatial coordinates only in a narrow band about the zero level set, for which fast methods exist to extract.\(^\text{27}\)

### 3.3 Data preparation

We use the data provided in the LIDC dataset\(^\text{13}\) for our experiments. The data undergoes a number of pre-processing steps, which we first describe briefly and then in more detail in the subsequent subsections. First, the data is analyzed to obtain only nodules where all four annotators agree upon the existence of a lung nodule at a particular location in the scan; this yields 896 nodules. Next, the radiologist annotation contours are converted to boolean-valued target volumes, and multiple annotations for the same nodule are consolidated into a single boolean volume. Afterward, for each of the 896 lung nodules selected, we standardize the image volumes containing the lung nodule as well as its associated ground-truth, boolean-valued volumes to have uniform voxel spacing of one millimeter because the CT scans in the LIDC data have been generated with different scanning devices and scanner parameters, resulting in different image volume resolutions. Lastly, we randomly partition this dataset of 896 lung nodules and respective ground-truth segmentations into subsets of size 672, 112, and 112 for training, validation, and testing, respectively.

#### 3.3.1 Gathering nodules with four annotators

Images of the lung nodules in the CT scan volumes from the LIDC were annotated by up to four radiologists, but the physical lung nodules lack a universal identifier. By “annotation”, we mean the sequence (i.e., the sequence through the slices of a particular CT scan) planar curves describing the boundary of a particular lung nodule as determined by a single radiologist. Symbolically, a lung nodule annotation can be written \( \mathcal{A} = (C_j, C_{j+1}, \ldots, C_{j+n}) \), where \( C_k \) is the curve describing the nodule boundary in image-slice \( k \) of the CT volume. In our experiments, we only use lung nodules that have been annotated by all four radiologists, and thus, we begin by estimating when multiple annotations refer to the same physical nodule in an image. We accomplish this by first defining a distance function, \( d \), that describes the nearness of two nodule annotations, \( \mathcal{A}_i \) and \( \mathcal{A}_j \). Next, for a given scan, we compute a distance matrix, \( D \), where the \((i, j)\) entry is \( d(\mathcal{A}_i, \mathcal{A}_j) \) (i.e., the distance
from annotation \(i\) to annotation \(j\), thus providing the pair-wise distances between all annotations in the scan. Two annotations are said to be adjacent when \(D_{ij} \leq \tau\), where \(\tau\) is a threshold parameter. The value of \(\tau\) is initialized to be equal to the slice thickness (in millimeters) of the scan, which is a parameter that can be found in the DICOM image data in the LIDC dataset. Nodule annotations are said to refer to the same physical nodule in the scan when they belong to the same connected component of the adjacency graph, which is formed by thresholding the pair-wise distance matrix \(D\) by \(\tau\). If afterwards there are annotation groupings with size greater than four nodules (i.e., greater than the number of annotating radiologists), we reduce the threshold parameter \(\tau\) by a multiplicative factor, and we repeat the process. In our work, we find that the distance function, \(d\), between two annotations that takes the minimum over all pairwise 3D distances between the coordinates of the two annotations works well, which is confirmed visually. From this process, we obtain 896 lung nodules, each having annotations by exactly four radiologists. This approach for clustering annotations is implemented in the pylidc\(^1\) Python package.

3.3.2 Volume interpolation

The pixel and slice spacing (i.e., the within- and between-slice scan resolutions, respectively) varies from among scan in the LIDC dataset, and to normalize this, we construct bounding boxes about each nodule, which we then interpolate to have uniform voxel spacing across all scans. For each lung nodule, of the 896 obtained with four annotations (discussed in the previous section), a common reference frame (i.e., a bounding box in the image volume) for the four associated annotations is formed. In this common reference frame, we convert each annotation into a boolean-valued volume that is one inside each annotating contour and zero outside, using the ray-casting method implemented in the matplotlib\(^2\) Python package. Next, we perform a tri-linear interpolation on both the image and the four associated boolean-valued volumes so that the resulting volumes have equal, one-millimeter spacing between voxels. The volumes are interpolated so that each volume is 70\(^3\) cubic millimeters, which was chosen to account for the nodule with largest observed diameter of 60 millimeters and to leave sufficient padding of non-nodule voxels about every nodule. Thus, the interpolated volumes are of dimensions, \(71 \times 71 \times 71\).

3.3.3 Consolidation of multiple ground truths and final pre-processing steps

For each nodule, we consolidate the four ground-truth segmentations (i.e., the boolean-valued indicator volumes of the lung nodule), \(B^k, k = 1, 2, 3, 4\) into a single a single ground-truth segmentation, \(B^*\), by computing, \(B^* = \arg\max\B_{B \in \Omega} \left\{\frac{1}{4} \sum_{k=1}^{4} J(B, B^k)\right\}, \quad \Omega = \{0, 1\}^{71 \times 71 \times 71}\) where \(J(\cdot, \cdot)\) is the Jaccard overlap function. The quantity \(B^*\) is sometimes referred to as the “Jaccard median” and is the best consolidation of the four annotations in the sense that it agrees most with all four annotations under the Jaccard overlap measure. Although we use the computed Jaccard median for our experiments, we observe that the typically-used 50% consensus consolidation (where the combined segmentation is equal to one where 50% or greater agreement occurs in the nodule’s segmentations and is zero otherwise) is often in high agreement with the Jaccard median. For example, in our case, \(J(B_{50}, B^*) \approx 0.96\) on average, where \(B_{50}\) denotes the 50% consolidation. This, combined with its simplicity (whereas the Jaccard median requires numerically solving a constrained optimization problem), justifies the use of the 50% consensus consolidation method that is often used in other works that use the LIDC dataset.

\(^1\)https://github.com/notmatthancock/pylidc
As the final pre-processing step, we standardize each image volume by subtracting off each respective mean and dividing by each respective standard deviation. This process results in datasets, \((M(k), B(k))_{k=1}^{N}\), of image and segmentation pairs, where \(N = 672\) for the training dataset and \(N = 112\) for the validation and testing datasets.

### 3.4 Initialization

In devising an initialization procedure, we use the observation that lung nodules are often, but not always, dense (and thus appearing relatively brighter than their surrounding in CT image slices), and approximately spherical in shape. These assumptions about the shape and appearance of the nodules are frequently violated, but nevertheless, the goal in initialization is to only to provide reasonable guess and to allow the LSML algorithm to improve upon it.

The initialization process uses local thresholding, connected component analysis, and a “radius trimming” post-processing technique. An example is shown in Fig. 3. Local thresholding of the image yields regions in the image that have similar image intensity values. Connected component analysis removes extraneous binary components obtained by thresholding, except the component that is closest to the seed point. Finally, the radius trimming technique serves to remove any parts of the binary component that extend beyond a specified distance from the center of the volume. This initialization process involves two free parameters, a smoothing factor \(\sigma\) and a radius percentile value \(p_r\), that are calibrated by performing a grid search over training data.

In more detail, the procedure begins by convolving a Gaussian smoothing kernel \(G_\sigma\) (with parameter \(\sigma\)) with the image, and the image is thresholded by sending values that are larger than the smoothed value to one and those that are below to zero. In other words, we transform the image to a boolean-valued image by setting those voxels to one that are greater than the average value in the neighborhood of surrounding voxels (where the neighborhood size is implied by the smoothing factor \(\sigma\)). Afterwards, we determine the connected component of this binary image that is closest to the center point of the volume. All pixel values in the binary image that are not in this connected component are set to zero.

Next, in spherical coordinates with the center of the image as the origin, we sample azimuth and zenith angles uniformly and determine the corresponding radius for given azimuth and zenith angles, defined as the distance until the first voxel where a ray emanating from the center of the volume meets a value of zero in the boolean volume. After sampling many azimuth and zenith angles, we compute the \(p_r\) percentile of radii observed. Any ray with radius extending beyond the \(p_r\) percentile is trimmed (see Figure 3b, left in blue), i.e., the thresholded image is set to zero outside a sphere centered in the image volume with radius corresponding to the \(p_r\) percentile of the radii observed. If after trimming the radii, there are multiple connected components, we choose the one closest to the seed point and set the others to zero. The initial values of \(u^0\) are set to +1 where the final boolean initialization is 1 and −1 where the boolean initialization is 0.

The free parameters \(\sigma\) and \(p_r\) are determined through a grid search procedure over the training data. Specifically, we search over the parameter values, \(\sigma \in \{1, 2, \ldots, 7\}\) and \(p_r \in \{50, 55, \ldots, 80\}\). For each parameter combination in the Cartesian product of these sets of values, we compare this segmentation from our initialization procedure against the ground truth segmentation under the Jaccard overlap score. The parameter combination with the highest average overlap score over the training data is used, which we determine to be \(\sigma = 4\) and \(p_r = 70\).
(a) **Left:** The center (35th) slice of the image volume is shown, as well as the center point of the image, indicated by the ‘seed’ label. **Right:** The image is thresholded by comparing the value at each voxel to the weighted average of the neighboring values. A connected component analysis for the component at the seed point reduces the number of components.

(b) **Left:** Radii are sampled by sampling azimuth and zenith angles, and computing the distance until the background value (i.e., a value of zero in the binary image volume) is first reached from a ray beginning at the seed point. Radii beyond the 70th percentile of the observed radii are trimmed to obtain the initialization shown in the blue curve in the center slice. The red curve shows the ground truth in this slice, for comparison. **Right:** The initialization volume is shown in red and blue hues, where the color on the surface indicate the distance from the ground truth. Red indicates a distance of zero and blue indicates the maximal distance of 7 voxels. The color-to-distance encoding is shown in the color bar in the figure. The center (35th) slice of the image volume is shown below the surface.

Fig 3: Example of level set initialization procedure for lung nodule images.
Table 2: Feature Map 1: image features are computed at fine ($\sigma = 0$) and coarse ($\sigma = 3$) scales. The symbols $\Omega = \{x : u > 0\}$ and $\partial \Omega = \{x : u = 0\}$ are used.

| Feature                     | Local | Global |
|-----------------------------|-------|--------|
| Volume $= |\Omega|$ | ✓      |        |
| Surface area $= |\partial \Omega|$ | ✓      |        |
| Isoperimetric ratio $= \frac{36\pi |\Omega|^2}{|\partial \Omega|^3}$ | ✓      |        |
| Moments of $\Omega$, order $= 1, 2$ | ✓      |        |
| Distance from $(i, j, k)$ to center of mass | ✓      | ✓      |
| Image average in $\Omega$ | ✓      |        |
| Image variability in $\Omega$ | ✓      |        |
| Image value at $(i, j, k)$ | ✓      |        |
| Image edge at $(i, j, k)$ | ✓      |        |

3.5 Features used

The first feature set that we apply to the lung nodule image segmentation problem, which we call Feature Map 1, is enumerated in Tab. 2. It consists of a number of simple and generic, global and local, shape and image features. These features serve as a baseline against the next feature set that extends Feature Map 1. All image features are computed at two scales, a fine scale ($\sigma = 0$) and a coarse scale ($\sigma = 3$), so that Feature Map 1 comprises a total of 18 features.

Table 3: Feature Map 2: image features are computed at fine ($\sigma = 0$) and coarse ($\sigma = 3$) scales. The symbols $\Omega = \{x : u > 0\}$ and $\partial \Omega = \{x : u = 0\}$ are used.

| Feature                               | Local | Global |
|---------------------------------------|-------|--------|
| All Feature Map 1 features (see Tab. 2) | ✓      | ✓      |
| Image average over $\partial \Omega$  | ✓      |        |
| Distance to center of mass average    | ✓      |        |
| Distance to center of mass variability| ✓      |        |
| Distance to center of mass maximum    | ✓      |        |
| Slice areas                           | ✓      | ✓      |
| Slice areas absolute change           | ✓      | ✓      |
| Image samples along normal            | ✓      | ✓      |
| Image samples along ray to center of mass | ✓      | ✓      |

The second feature map, Feature Map 2, is enumerated in Tab. 3. It extends Feature Map 1 by including additional local image and shape features. Many of the features in Feature Map 2 include both local and global aspects in their computation. These additional global-local (or ‘glocal’) features are local in the sense that they require the local voxel coordinate $(i, j, k)$, but use previously computed global features such as the “center of mass” in their computation. Feature Map 2 includes a total of 109 features, where all image features are computed at both fine and coarse scales ($\sigma \in \{0, 3\}$). The “distance to center of mass” statistics features in Feature Map 2 (i.e., average, variability, and maximum) supplement the “distance to center of mass” feature from Feature Map 1 with more global context. The “slice areas” and “slice areas absolute change” features
are computed by calculating the areas of each slice through $\Omega$ in the three axes directions. These provide a more localized version of the purely global volume feature in Feature Map 1, and in addition, they attempt to help enforce a slice-to-slice continuity. The “image samples along normal” and “image samples along ray to center of mass” features, illustrated in Fig. 4, provide context of the image along two lines emanating from a given coordinate $x_0 = (i, j, k)$. Image values are sampled along two lines: (1) in the direction of the unit normal through $x_0$, and (2) in the direction of the line segment connecting the center of mass and $x_0$. For each of these lines, 10 samples are taken in the direction inward and outward from $x_0$.

4 Results

Table 4: Results for Feature Map 1 and Feature Map 2. $n^*$ is the iteration number when the LSML algorithm terminates, and $\bar{J}$ is the average Jaccard overlap score over the testing dataset.

|               | $n^*$ | $\bar{J}$       |
|---------------|-------|-----------------|
| Feature Map 1 | 45    | 0.6951 (±0.1119)|
| Feature Map 2 | 47    | 0.7185 (±0.1114)|

In Tab. 4 we report the average Jaccard overlap scores over the testing set at the optimal iteration from the validation dataset. In Fig. 5a, we plot the average Jaccard overlap score over the testing dataset against the iteration number for Feature Map 1 (shown in solid black) and Feature Map 2 (shown in dashed black). At initialization, the average Jaccard overlap score is 0.6484 (±0.1119). Using Feature Map 1, the LSML algorithm terminates after 45 iterations, obtaining a final average Jaccard overlap score of 0.6951 (±0.1127), an increase of 7.2% relative to the average overlap score at initialization, whereas with Feature Map 2, the algorithm terminates after 47 iterations, and obtains a final average overlap score of 0.7185 (±0.1114), a 10.8% relative increase over the average score at initialization. Feature Map 2 produces a 3.4% relative increase in overlap score over the overlap score produced under Feature Map 1, which is a modest, but significant ($p = 2 \times 10^{-6}$) increase, indicating the usefulness of the ‘glocal’ features that are included in Feature Map 2. We also observe from Figure 5a that the score increases more rapidly during the earlier iterations than the latter, indicating that the more substantial changes in the segmentation
(a) Average Jaccard overlap scores over the testing dataset against iteration number for Feature Maps 1 and 2.

(b) Box plots of the distribution in Jaccard overlap score for Feature Map 1 and Feature Map 2.

Fig 5: Results for the LSML method applied to lung nodule image volumes from CT scans in the LIDC dataset.

shape (in the sense of those that produce relatively larger increases in the overlap score) occur in the early iterations, whereas the latter iterations serve in making only small refinements. In Figure 5b, the distributions of overlap scores for the Feature Maps 1 and 2 are given on the left and right, respectively. The median overlap score for Feature Map 1 is 0.7115 (above the mean overlap of 0.6951), and the 25th percentile to the 75th percentile stretches from 0.6596 to 0.7650, having a range of 0.1054. For Feature Map 2, the median overlap score of 0.7356 (above the mean overlap of 0.7185) with the 25th percentile to the 75th percentile ranging from 0.6793 to 0.7918, having a range of 0.1125. The large range in overlap score indicates the difficulty of the lung nodule segmentation problem compared to the synthetic dataset examples discussed in previous sections.

The maximal overlap scores observed for Feature Maps 1 and 2 are 0.8593 and 0.8906, respectively. The lung nodule with highest overlap score under Feature Map 2 is shown in Figure 6. This nodule achieves the second highest overlap score (of 0.8573) under Feature Map 1 and is an isolated nodule with a well-defined boundary. Both feature maps yield segmentations that capture the true boundary very well, with the most notable difference being in the 32nd slice, where the true boundary includes slightly more of the region of the nodule that is in close proximity to the vasculature in the posterior direction (i.e., moving downward from the lung nodule in the image).

Two outliers are observed (see Figure 5b) under both feature maps, the lowest overlap score having an overlap score of 0.0582 and 0.0732, for Feature Maps 1 and 2, respectively, which resulted from a poorly initialized segmentation of a juxta-pleural lung nodule in a region near the bottom of the lung with the nearby lung wall and organs having image values very close to those of the nodule, as can be seen in Figure 7. This case is an outlier, and there are many juxta-pleural nodule cases (as well as other nodule anatomical location and density categories) where substantial
Fig 6: The lung nodule with maximal overlap score under Feature Map 2 and second highest overlap score under Feature Map 1.

(a) The nodule (at the tip of the red arrow) is shown in the context of the entire slice of the CT scan to which it belongs.

(b) The center (35th) slice through the lung nodule image volume is shown with the true segmentation in red along final iterations of the LSML method under Feature Maps 1 and 2, shown in blue and yellow, respectively.

Fig 7: The lung nodule with minimal overlap score for both Feature Maps 1 and 2.

Improvements are observed from Feature Map 1 to Feature Map 2.

In Figure 8, we show the center (35th) slice through the image volume of each of the 112 lung nodules in the testing dataset, where the red curve represents the contour given by the slice through the ground-truth segmentation surface and the blue curve represents the contour given by the slice through the approximate segmentation surface given by the zero level set in the LSML method using Feature Map 2. The LSML method performs well in a variety of contexts, including many juxta-pleural nodules (e.g., row five, column two; or, row three, column eight), nodules with cavities (e.g., row one, column eleven; or, row seven, column ten), non-solid nodules (e.g., row four, column nine), irregularly-shaped nodules (e.g., row nine, column three), spiculated nodules.
(e.g., row ten, column five; or, row three, column nine), as well as the other nodule types shown. These visual results, together with the quantitative results discussed previously, demonstrate the effectiveness of the LSML method applied to the lung nodule image segmentation problem in CT image volumes.

Fig 8: The center (35th) slice of each of the 112 lung nodules from the testing dataset. The red contour is produced by slicing the ground-truth segmentation through its center slice, and the blue contour is the approximate segmentation obtained by taking the center slice through the zero level surface produced by the LSML method using Feature Map 2.
5 Conclusions

Lung nodule segmentation is a core component of the lung CAD pipeline, and accurate nodule segmentation poses unique challenges. The LSML method, a natural and direct machine learning extension of the level set image segmentation method, achieves an average Jaccard overlap of 0.7185 (±0.1127), which is comparable to the current state-of-the-art for lung nodule image segmentation. The mortality rate of lung cancer is large, and lung CAD methods, if robustly validated in clinical settings, carry the potential to aid physicians towards increased the survival rate of those afflicted. Accurate lung nodule segmentation is work towards this goal.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

References

1. R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2017,” CA: A Cancer Journal for Clinicians 67, 7–30 (2017).
2. M. F. McNitt-Gray, S. G. Armato, C. R. Meyer, et al., “The Lung Image Database Consortium (LIDC) Data Collection Process for Nodule Detection and Annotation,” Academic Radiology 14, 1464–1474 (2007).
3. J. Erasmus, J. Connolly, H. McAdams, et al., “Solitary pulmonary nodules: part I. Morphologic evaluation for differentiation of benign and malignant lesions,” Radiographics (2000).
4. C. I. Henschke, D. F. Yankelevitz, D. M. Libby, et al., “Survival of Patients with Stage I Lung Cancer Detected on CT Screening,” New England Journal of Medicine 355(17) (2006).
5. R. Malladi and J. Sethian, “Shape modeling with front propagation: A level set approach,” IEEE Transactions on Pattern Analysis and Machine Intelligence 17(2) (1995).
6. M. Kass, A. Witken, and D. Terzopoulos, “Snakes: active contour models,” International Journal of Computer Vision 1(4), 321–331 (1988).
7. S. Osher and J. A. Sethian, “Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations,” Journal of computational physics 79(1), 12–49 (1988).
8. M. E. Leventon, O. Faugeras, W. E. L. Grimson, et al., “Level set based segmentation with intensity and curvature priors,” in Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on, 4–11, IEEE (2000).
9. A. Tsai, A. Yezzi, W. Wells, et al., “A shape-based approach to the segmentation of medical imagery using level sets,” IEEE Transactions on Medical Imaging 22, 137–154 (2003).
10. D. Cremers, M. Rousson, and R. Deriche, “A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape,” International Journal of Computer Vision 72, 195–215 (2007).
11. D. Cremers and M. Rousson, “Efficient kernel density estimation of shape and intensity priors for level set segmentation,” in Deformable Models, 447–460, Springer (2007).
12. B. Van Ginneken, “Supervised probabilistic segmentation of pulmonary nodules in CT scans,” in Medical Image Computing and Computer-Assisted InterventionMICCAI 2006, 912–919, Springer (2006).
13 S. G. Armato III, G. McLennan, L. Bidaut, et al., “The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans,” *Medical physics* **38**(2), 915–931 (2011).

14 S. Wang, M. Zhou, Z. Liu, et al., “Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation,” *Medical Image Analysis* **40**, 172–183 (2017).

15 R. Tachibana and S. Kido, “Automatic segmentation of pulmonary nodules on CT images by use of NCI lung image database consortium,” (2006).

16 Q. Wang, E. Song, R. Jin, et al., “Segmentation of Lung Nodules in Computed Tomography Images Using Dynamic Programming and Multidirection Fusion Techniques1,” *Academic Radiology* **16**, 678–688 (2009).

17 T. Messay, R. C. Hardie, and S. K. Rogers, “A new computationally efficient CAD system for pulmonary nodule detection in CT imagery,” *Medical Image Analysis* **14**, 390–406 (2010).

18 T. Messay, R. C. Hardie, and T. R. Tuinstra, “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,” *Medical Image Analysis* **22**, 48–62 (2015).

19 T. Kubota, A. K. Jerebko, M. Dewan, et al., “Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models,” *Medical Image Analysis* **15**, 133–154 (2011).

20 B. C. Lassen, C. Jacobs, J.-M. Kuhnigk, et al., “Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans,” *Physics in Medicine and Biology* **60**, 1307–1323 (2015).

21 Y. Tan, L. H. Schwartz, and B. Zhao, “Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field,” *Medical physics* **40**(4), 043502 (2013).

22 J. S. Schildkraut, S. Chen, M. Heath, et al., “Level-set segmentation of pulmonary nodules in radiographs using a CT prior,” *Proc. SPIE Med. Imag., Image Process.* **7259**, 72593B–1 (2009).

23 A. A. Farag, H. E. A. El Munim, J. H. Graham, et al., “A Novel Approach for Lung Nodules Segmentation in Chest CT Using Level Sets,” *IEEE Transactions on Image Processing* **22**, 5202–5213 (2013).

24 M. M. Farhangi, H. Frigui, A. Seow, et al., “3d Active Contour Segmentation Based on Sparse Linear Combination of Training Shapes (SCoTS),” *IEEE Transactions on Medical Imaging*, 1–1 (2017).

25 T. F. Chan and L. A. Vese, “Active contours without edges,” *Image processing, IEEE transactions on* **10**(2), 266–277 (2001).

26 D. E. Breen and R. T. Whitaker, “A level-set approach for the metamorphosis of solid models,” *IEEE Transactions on Visualization and Computer Graphics* **7**(2), 173–192 (2001).

27 D. Adalsteinsson and J. A. Sethian, “A fast level set method for propagating interfaces,” *Journal of Computational Physics* **118**(2), 269–277 (1995).

28 J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing In Science & Engineering* **9**(3), 90–95 (2007).
“American cancer society cancer facts and figures 2017.” https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html. Accessed: Jul. 2017.