Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks

Xia Huang\textsuperscript{a}, Wenging Sun\textsuperscript{b}, Tzu-Liang (Bill) Tseng\textsuperscript{c}, Chunqiang Li\textsuperscript{d,∗,1}, Wei Qian\textsuperscript{b,∗,2}

\textsuperscript{a} Department of Metallurgical, Materials, and Biomedical Engineering, University of Texas at El Paso, El Paso, TX, 79968, United States
\textsuperscript{b} Department of Electrical and Computer Engineering, University of Texas at El Paso, El Paso, TX, 79968, United States
\textsuperscript{c} Department of Industrial, Manufacturing, and Systems Engineering, University of Texas at El Paso, El Paso, TX, 79968, United States
\textsuperscript{d} Department of Physics, University of Texas at El Paso, El Paso, TX, 79968, United States

\textbf{A R T I C L E   I N F O}

Article history:
Received 20 September 2018
Received in revised form 9 January 2019
Accepted 18 February 2019

Keywords:
Computer aided diagnosis
Pulmonary nodule detection and segmentation
Convolutional neural networks
Faster regional-CNN
Fully convolutional neural network (FCN)

\textbf{A B S T R A C T}

Deep learning techniques have been extensively used in computerized pulmonary nodule analysis in recent years. Many reported studies still utilized hybrid methods for diagnosis, in which convolutional neural networks (CNNs) are used only as one part of the pipeline, and the whole system still needs either traditional image processing modules or human intervention to obtain final results. In this paper, we introduced a fast and fully-automated end-to-end system that can efficiently segment precise lung nodule contours from raw thoracic CT scans. Our proposed system has four major modules: candidate nodule detection with Faster regional-CNN (R-CNN), candidate merging, false positive (FP) reduction with CNN, and nodule segmentation with customized fully convolutional neural network (FCN). The entire system has no human interaction or database specific design. The average runtime is about 16 s per scan on a standard workstation. The nodule detection accuracy is 91.4% and 94.6% with an average of 1 and 4 false positives (FPs) per scan. The average dice coefficient of nodule segmentation compared to the groundtruth is 0.793.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Worldwide, lung cancer has been having the leading mortality rate of cancer deaths in both males and females for decades with 1.2 million global deaths a year (Fitzmaurice et al., 2017). Due to the inconsiderable symptoms, the majority of lung cancer cases are diagnosed at distant stages with only 4% five-year survival rate (Siegel et al., 2018). Early detection of suspicious pulmonary nodules is crucial to improve the life quality of lung cancer patients. Currently, computed tomography (CT) is considered the best and most widely used imaging modality for early detection and analysis of lung nodules. However, because of the complicated morphological and anatomical appearance of nodules, the nodule identification would be largely dependent on the skill, experience, and vigor of the radiologists (Winkels and Cohen, 2018). After identifying the nodule, precise segmentation is significant for clinical measurements (such as diameter and volume), which objectively provides repeatability of diagnosis and consistency of image interpretation (Liu et al., 2018). Therefore, a fast and fully-automated computer aided detection (CAD) system on nodule detection and segmentation with limited number of false positives (FPs) will dramatically decrease the workload of radiologists as well as the cost of treatment.

For nodule detection many published works proposed a two-stage system, which includes a candidate screening step to rapidly extract nodule regions from pulmonary parenchyma and remove other structures, and a false positive (FP) reduction step to massively eliminate FP candidates from the detected ones until reaching clinically acceptable performance. Ge et al. (2005) used adaptive weighted k-means clustering to segment suspicious candidates, and reduced FPs by combing 3D gradient field and ellipsoid features with a linear discriminant analysis (LDA) classifier. Li et al. (2008) added a multiscale selective filter to enhance nodule and
and 26 remove CNNs, detection on the 2016 prevalent a features et presentations five of traditional restrictions, and the tedious experiment-based parameter-adjustment restricts the applications onto real clinical trials. Moreover, the suboptimal segmentation of lung parenchyma negatively impacts the candidate screening especially for juxtapleural nodules (Dai et al., 2015). Second, because of the serial algorithm structure with many subcomponents, the long computation time stands out as another demerit. Therefore, a simpler and more independent nodules detection framework is urgently desired.

Because of the critical clinical value in nodules segmentation, a growing numbers of pulmonary nodule segmentation algorithms have been proposed in literature. They can be roughly categorized into four types: 1) Threshold based methods (Reeves et al., 2006; Magalhães Barros Netto et al., 2012; Tachibana and Kido, 2006 Xia et al., 2016). For instance, Tachibana and Kido (2006) designed a coarse-to-fine scheme that consists of a rough segmentation step using multiple fixed thresholds to roughly identify the nodes region and a precise segmentation step using a watershed-based algorithm to remove the unnecessary structures attached to them. 2) Morphology based methods (Kubota et al., 2011; Dehmeshki et al., 2008; Vijaya Kishore and Satyanarayana, 2013 Lassen et al., 2015). For example, Dehmeshki et al. (2008) presented an efficient sphericity-oriented region growing algorithm applied on the fuzzy connectivity mask created by a connectivity region growing technique with only one single seed point provided by the user. 3) Statistical model based method (Wang et al., 2009; Dong and Peng, 2014; Tan et al., 2013; Mao et al., 2018). Tan et al. (2013) utilized a hybrid algorithm combining marker-controlled watershed, geometric active contours as well as Markov random field (MRF). Similar to the method Tachibana et al. proposed, they imposed watershed method to generate an initial surface of nodules, followed by the refinement of active contours. And MRF optimally estimates the texture distribution of ground glass opacity, so that it improves the segmentation accuracy for this portion. 4) Clustering methods based on traditional machine learning (Van Ginneken, 2006; Tuinstra, 2008; Messay et al., 2015). Messay et al. (2015) proposed a selective regression neural network (RNN) based algorithm with both fully-automated and semi-automated options. The feature learning process using RNN can automate the parameter setting for each nodule based on the learned features. However, the majority of the aforementioned methods perform well only on specific type of nodules (e.g. solitary pulmonary nodule) or on relatively small size of dataset, which cannot satisfy the variety and complexity of pulmonary nodules. In addition, most of the methods still need human interventions, which largely undermine the purpose of CAD systems. Finally, in order to achieve optimal performance, most of the techniques require massive iterations and parameter tunings, which to a large extent, slow down the overall computation process. Recently, the success of semantic segmentation in computer vision field based on fully convolutional neural networks (FCNs) (Long et al., 2015; Wang et al., 2017; Chen et al., 2018; Lekić and Babić, 2018 Yu et al., 2018) attracts some researchers to concentrate on the application of pulmonary nodule segmentation. Wu et al. (2018) firstly deployed an interpretable and multi-task CNNs model to segment and classify pulmonary nodules by feeding 3D patches and achieved the state-of-the-art performance. However, most of current deep learning based algorithms still rely on several preprocessing steps such as lung parenchyma segmentation, which decrease the level of automation.

In this paper, we propose a CNNs based algorithm to automatically detect and segment pulmonary nodules with very limited
number of FPs. Comparing with methods using the original nodule patch only, our method improves the area under the curve (AUC) malignancy prediction by 4% with the combination of the original nodule patch, the segmented nodule patch highlighting the shape, and the gradient of nodule patch highlighting the texture with accurate boundary information provided (Sun et al., 2017b).

To our best knowledge, this is one of the first works that exploit pure 2D CNNs based algorithm for pulmonary nodule segmentation from the raw CT scan without any manual settings. The output is the corresponding nodule masks with high overlapping ratio compared to radiologists’ markings. This study is the follow up study of our previous researches (Qian et al., 1993; DeVore et al., 1995; Sun et al., 2004; Zhang et al., 2007; Ye et al., 2013; Sun et al., 2017a,b, 2016, 2017c, 2017d, 2017e). The reason of using 2D axial slice instead of 3D volume in our system is three-fold. First, a powerful deep learning model relies on large amount of training data. However, due to the stressful workload of manual nodule identification, the available thoracic CT scans with high quality groundtruth annotations are still insufficient to train a robust and discriminative deep learning network. Since a nodule may appear in several neighboring 2D slices of a single CT scan, using 2D instead of 3D slices will naturally augment the size of training set without manual data augmentation process. Second, 3D deep learning models will exponentially add computation complexity compared to 2D models. Therefore, some models require image down sampling or cropping (Jin et al., 2018) to compensate the massive memory consumption. This requirement, to some extent, restricts the feasibility of 3D models on common workstation with limited GPU resources. Third, transfer learning (Weiss et al., 2016; Long et al., 2016) can be performed based on some fine-tuned 2D models such as VGG16 (Simonyan and Zisserman, 2015), ResNet-50, and ResNet-101 (He et al., 2016), significantly accelerates the convergence of training, boosts the performance, and cuts down the computation complexity. The remaining paper is organized as follows. The data and methodologies will be described in Section 2. Section 3 reports the experimental results with qualitative and quantitative performance evaluation, execution performance analysis, and comparison between our proposed method and some state-of-the-art approaches to underline the significance of our study. Conclusions and some potential improvements are stated in Section 4.

2. Materials and methods

2.1. Dataset

Similar to most of aforementioned literatures, we train and evaluate our fully-automated system on a large publicly available dataset, organized by Lung Image Database Consortium image collection (LIDC-IDRI) (Armato et al., 2015, 2011; Clark et al., 2013) with 1018 scans from seven academic centers and eight medical imaging companies. The slice thickness of these scans is ranging from 0.6 mm to 5.0 mm. For each scan, four experienced radiologists performed two-phase nodule assessment and recorded the detailed nodule information such as boundary coordinates, malignancy level etc. into an XML file. In this data set three types of lesions are included: non-nodule, small nodule (<3 mm), and large nodule (≥3 mm). According to some clinical recommendations (Aberle et al., 2011), we only consider large nodules (≥3 mm) in our study. The pre-filtering strategy proposed in the Lung Nodule Analysis 2016 (LUNA16) challenge only selected nodules with the consensus of at least three out of four radiologists with slice thickness no more than 2.5 mm. As a consequence, 888 CT scans with 1186 nodules are involved.

2.2. Methodology

The framework of our proposed nodule segmentation algorithm is shown in Fig. 1. It can be simply divided into four main components: 1) 2D Faster R-CNN based candidate detection to rapidly locate pulmonary nodule patches; 2) merging overlapping candidates by combining 2D patches with close Euclidean distances; 3) traditional three-layer 2D CNN based FP reduction to further eliminate FPs; and 4) modified FCNs based nodule segmentation to precisely segment the initial nodule mask. The geometric centers of detected nodules will guide the system to refine the segmentation and output the final nodule segmentation result. The more detailed explanation of each component is provided below.

2.2.1. Preprocessing

In order to reduce memory consumption, we assigned -1000 HU (air) as lower bound and +3000 HU (bone) as upper bound, then applied a linear mapping to convert the original signed 16-bit CT volume scans into 8-bit intensity values in the range of 0-255. This is the only preprocessing work in our proposed approach.

2.2.2. Faster R-CNN based nodule detection

The Faster R-CNN model evolved from Fast R-CNN (Girshick, 2015), which is mainly composed of a region proposal network (RPN) to propose potential regions of objects and an object detection network (ODN) to classify the region proposals from RPN. The new contribution of Faster R-CNN is that RPN and ODN can share the same convolutional layers, which enabled a more unified system to run at near real-time frame rates on natural images without performance loss (Ren et al., 2017). Based on our experiments, it is
fully convertible to be applied on nodule detection after conducting several modifications. The network architecture is shown in Fig. 2. To accelerate training convergence and save computation workload, we reused the weights of first five groups of convolutional and pooling layers with a total of thirteen consecutive convolutional (Conv) and five pooling layers from a pretrained VGG16 model, and the calculated feature maps are shared with both RPN and ODN training. For RPN, we utilized a small network by sliding a 3 × 3 window at a time over the shared feature space to convert this feature map to a 512-dimension feature vector followed by a ReLU layer (Nair and Hinton, 2010). We implement the aforementioned small network by a 3 × 3 convolutional layer. After two sibling 1 × 1 convolutional layers, an RPN regression (rpn_reg) layer outputs the bounding box coordinates of each proposal, and an RPN classification layer (rpn_clz) estimates the probability of the proposal being a nodule. The design of anchors ensures the capability to parallelly predict multiple nodule proposals at each sliding window location. Different from common objects in natural images with big sizes and elongated shapes, the nodules are with relatively small and square boundaries. Therefore, we used a fixed scale ratio (1:1) and removed the other scale ratios in original RPN design, and implemented seven anchors with ascending common differences: 4 × 4, 6 × 6, 10 × 10, 16 × 16, 24 × 24, 36 × 36, 52 × 52 to fit the size variations of nodules. Because of the sparse distribution of nodules, we also adopt non-maximum suppression (NMS) based on the scores of rpn_clz with intersection over union (IoU) threshold of 0.7 between groundtruth and RPN proposals. NMS massively reduces the number of proposals and also potentially improves the detection accuracy, since high-density proposals may cause RPN to identify many surrounding regions that partially overlapped with true nodules (please see Fig. 3 as an example). The learnt proposals from RPN are fed into ODN for further classifications.

Taking the proposals predicted by RPN, the ODN is involved to serve as a binary classifier to determine nodule regions using the architecture of traditional CNNs. A ROI pooling layer is imposed to map each proposal to a smaller feature map by implementing max pooling operation of the values in a fixed 7 × 7 sub-window. Then two 4096-way fully-connected layers (FC1 and FC2) are conducted to produce a lower dimension feature vector, followed by two independent Softmax layers to output bounding boxes and probability scores of predicted nodules (box_reg and box_clz). In Faster R-CNN, RPN and ODN are mutually finetuned by adopting a pragmatic four-step alternating training procedure (Ren et al., 2017). As such, we achieved a unified network with sharing convolutional layers for RPN and ODN and the loss function for a single batch of N images is defined as follows.

\[
\begin{align*}
\mathcal{L}_r &= \frac{1}{N} \left[ \frac{1}{N_{tr}} \sum_i \mathcal{L}_r \left( t_i, t_i^* \right) + \frac{1}{N_{tr}} \sum_i \mathcal{L}_r \left( p_i, p_i^* \right) \right] \\
&+ \frac{1}{N_{tr}} \sum_i \mathcal{L}_r \left( t_i, t_i^* \right) + \frac{1}{N_{tr}} \sum_i \mathcal{L}_r \left( p_i, p_i^* \right) \\
\mathcal{L}_r \left( t, t^* \right) &= R(t - t^*) \\
\mathcal{L}_r \left( p, p^* \right) &= -\log[pp^* + (1 - p)[1 - p^*]]
\end{align*}
\]

where \(N_{tr}, N_{tr}, N_{br}, \) and \(N_{br}\) are numbers of inputs in rpn_reg, rpn_clz, box_reg, box_clz layers respectively, \(\mathcal{L}_r\) and \(\mathcal{L}_r\) represent loss associated the regression and classification layers, \(t_i\) represents the four coordinates of predicted nodule proposal, \(t_i^*\) is the coordinates of the corresponding groundtruth nodule, \(p_i\) and \(p_i^*\) denote the predicted and true probability of current anchor to be a nodule in RPN. Similarly, \(t_i, t_i^*, p_i, p_i^*\) implicit the same concepts in ODN. Besides, \(R\) is a robust loss function (smooth \(L_1\) loss) explained in (Girshick, 2015).

2.2.3. Merging of overlapping nodule candidates

After Faster R-CNN, most nodule candidates can be detected. However, a single nodule may appear in several slices, therefore it may have more than one candidate representing the same nodule especially for some nodules with bluryry edges. Based on some intuitive observations, these candidates will lie in close proximity to each other. Thus, a simple and computation-efficient merging operation is implemented by recursively combining candidates within five voxels of each other until no further merge is needed. This merging procedure ensures that a single nodule is identified within a single 2D slice rather than multiple slices alongside each other, dramatically decreases the number of unnecessary detections, and fastens the processing speed of the following FP reduction since the base number of candidates are lowered.

2.2.4. CNN based false positive reduction

With the nodule candidates extracted by Faster R-CNN, true nodule regions are successfully identified with small amount of
FPs. However, the existence of FPs still prohibits the use of CAD system in clinical practice. Considering the requirement of computation time and the advantages of Faster R-CNN component, a simple 2D CNN based classifier is sufficient to handle FP reduction task. Based on our previous work (Sun et al., 2017a) with the addition of multi-view nodule patches obtained from nine symmetrical planes presented in a previous work (Setio et al., 2016), an individual CNN with three pairs of convolution-pooling layers and one fully-connected layer is implemented. The loss is calculated by using the cross-entropy error, and weights are updated using mini-batches of 128 images. Then the testing is incorporated based on the Faster R-CNN initial candidate detection results. The kernel size for each convolutional layer is $5 \times 5$, $3 \times 3$, and $3 \times 3$ and the numbers of filters are 24, 48, 64, respectively. The input image size is $64 \times 64$ for both training and testing sets. After FP reduction step, the occurrence of FPs is largely decreased while maintaining a high detection sensitivity. Since the FP reduction network will be executed after Faster R-CNN initial candidate detection, different initial candidate FP levels with corresponding sensitivity may cause variations on the overall performance. In order to achieve the best performance, we empirically set three initial candidate FP levels (small, medium, large) and individually employ FP reduction on each FP level. We quantify the results in Section 3.

2.2.5. Modified FCN based nodule segmentation

FP reduction eliminated the most unlikely nodule candidates detected by Faster R-CNN. Among the remaining candidates, the 100 by 100 patches containing the nodule detection bounding boxes are created. Compared to FP reduction network that conducts a single imagewise two-class classification, more precise pixelwise classification is performed for segmentation purpose. Therefore, larger receptive field with richer background texture is needed. The modified deconvolutional neural network (Long et al., 2015) is used to generate the detailed nodule segmentation contour. In this experiment, the VGG16 is imposed as the backbone, and the weights of all the convolutional layers are initiated by ImageNet VGG16 pretrained model, while the weights of the later deconvolutional layers are randomized. In VGG16, there are five groups of convolutional layers altogether (Conv1 to Conv5), and each group contains a few consecutive convolutional layers. The original FCN used three convolutional layers (Conv3, Conv4 and Conv7) to generate the segmentation results. Since the lower convolutional layers have higher resolution, incorporating these layers should help with segmentation precision. In this study, two extra rounds of training are added to FCN training procedure to incorporate the first two convolutional layers (Conv1 and Conv2) to deconvolutional networks as well. The architecture of our modified FCN is shown in Fig. 4.

---

**Fig. 3.** Comparison of nodule candidate detection with and without non-maximum suppression (NMS) operation. (a) Nodule candidates without NMS operation. (b) Nodule candidates with NMS operation. (c) and (d) are zoom-in view of the region marked by yellow dashed square in (a) and (b). The predicted nodule candidates by Faster R-CNN are marked by red solid squares with the classification probability on top of them. Note that only the candidates with classification probability of nodules larger than 0.7 are displayed in this example. With NMS operation, our system successfully detects a ground glass opacity (GGO) nodule without partially overlapped duplicates (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).
To initiate the FCN, the ImageNet VGG16 weights are downloaded and a classification FCN is trained based on it. Each convolutional layer provides the local neighborhood information of the image, and each pooling layer down-samples the image by the stride of 2. To fine tune the VGG16, the third fully connected layer (fc8) is removed, and the output node is set to 2 since our dataset has two subsets: nodule and non-nodule. Then all the fully connected layers are converted to convolutional layers, and all the weights are preserved in the transformed model.

Then the fine-tuned VGG16 weights are used to initiate FCN32 s, and an extra deconvolutional layer connected to the last convolutional layer is initialized with random noise attached to the end. Then we up sample the deconvolutional layer result by 2 to generate the FCN32 s segmentation mask. The added deconvolutional layer and up sampling layer can be considered as a block, after FCN32 s being trained, another block is attached to the end to train FCN16 s. Because there is another up sample layer in FCN16 s compared to FCN32 s, the original deconvolutional results are twice as large. Different deconvolutional results are up sampled to the identical size as the original image, and concatenated together to generate the segmentation output. The whole process is repeated 5 times so the FCN2s output has the same size and resolution of the original image. All the deconvolutional layers are up sampled at the same size and concatenated together. Compared to original FCN8s, our model utilizes the lower level convolution results thus yielding higher resolutions and preserved the details of the nodule. Since segmentation is indeed a pixel wise classification task, we still use cross entropy loss as the loss function during the FCN training.

2.2.6. Postprocessing

The final segmentation results are obtained after the fusion of detected nodule centers and initial segmentation masks to remove non-nodule segmentations. This fusion operation only selects the corresponding segmented object with shortest Euclidean distances between detected nodule center and object centers in the initial segmentation mask.

3. Results and discussion

3.1. Experimental design

For the Faster R-CNN based nodule detection, we take the 888 filtered CT scans with 1186 high confidence pulmonary nodules from LUNA16 dataset and retrieve their original annotations with precise nodule boundary coordinates recorded in XML format from LIDC-IDLRI. To create the bounding square for each nodule, we calculate the smallest bounding rectangle and treat the longer side as the side of the result bounding square and record the four edge coordinates in groundtruth file. Because of the 2D network design, a total of 7909 axial nodule slices are generated. Leveraging from the ten patient-level subsets by LUNA16, we include nodule slices of subset0 to subset5 as training, subset6 to subset7 as validation, and subset8 to subset9 as testing. As a consequence, 534 scans with 5040 nodule slices, 178 scans with 1411 nodule slices, and 176 scans with 1458 nodules slices are assigned to training, validation, and testing set, respectively. For the training step, we adapt backpropagation and stochastic gradient descent (SGD) (LeCun et al., 1989) to train our end-to-end network. The minibatch size of images per iteration is set as 2 and the minibatch size of anchors per image is set as 10. For each image, based on the number of positive anchors (nodule), we randomly crop equal number of negative anchors (background) so that the sampled positive and negative anchors have the ratio of 1:1. Except for the first five groups of convolutional layers from VGG16, we randomly initialize all the other layers by drawing weights from a zero-mean Gaussian distribution with standard deviation 0.01. The learning rate is 0.001 for the first 10 epochs, and 0.0001 for the remaining with the momentum of 0.9 and a weight decay of 0.0005. Based on the aforementioned settings, the loss is stable after 20 epochs training. The total training time for Faster R-CNN is about 2.5 h. For the modified FCN based nodule segmentation, the learning rates for the five deconvolutional training are set to 10–12, 10–14, 10–16, 10–14, 10–15, respectively. The momentum is set to 0.99, weight decay is 0.0005, the iteration is set to 50,000 for each round of training. Besides the initialization round, we train FCN with five rounds manner, adding one fully connected layer each time. Each round of FCN training is about 5 h, and total training time for FCN is about 30 h.

All the training and testing of CNNs are implemented using Tensorflow (Abadi et al., 2015) on a standard PC with a single 6GB memory GPU GeForce GTX 980Ti.

3.2. Evaluation metrics

To evaluate nodule detection accuracy we calculate the detection sensitivity and the corresponding average FPs per scan for every prediction probability level, and the Free Receiver Operating Characteristic (FROC) is imposed to visualize the sensitivity versus
Table 1
The system performance and CPM score comparison of the proposed method and other state-of-the-art approaches. Note that “online” means models with online descriptions available on LUNA16 competition website: https://luna16.grand-challenge.org/Results/, **” represents models with limited details provided.

<table>
<thead>
<tr>
<th>Team</th>
<th>GPU</th>
<th>CNN Dimension</th>
<th>CPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAtech (PA_tech)online</td>
<td>Tesla K80 × 4</td>
<td>3D</td>
<td>0.951</td>
</tr>
<tr>
<td>JiaPanCAD (weiyjxie)online</td>
<td>NVIDIA Titan X × 2</td>
<td>3D</td>
<td>0.950</td>
</tr>
<tr>
<td>LUNA16FONOVACAD (zxp774747)online</td>
<td>NVIDIA M40 × 2</td>
<td>3D</td>
<td>0.947</td>
</tr>
<tr>
<td>iFLYTEK-MIG (yinbaoca)i</td>
<td>–</td>
<td>–</td>
<td>0.941</td>
</tr>
<tr>
<td>zhongliu_xie (zhongliu_xie)online</td>
<td>NVIDIA K80 × 2</td>
<td>3D</td>
<td>0.922</td>
</tr>
<tr>
<td>iDST-VC (chenxj1005)i</td>
<td>–</td>
<td>3D</td>
<td>0.897</td>
</tr>
<tr>
<td>dpjxd1 (dpjxd1) [Ding et al., 2017]</td>
<td>–</td>
<td>3D</td>
<td>0.891</td>
</tr>
<tr>
<td>CASED (CASED) (Jesson et al., 2017)</td>
<td>–</td>
<td>3D</td>
<td>0.887</td>
</tr>
<tr>
<td>3DCNN_NDET (lishaxuei)i</td>
<td>–</td>
<td>3D</td>
<td>0.882</td>
</tr>
<tr>
<td>Aidence (mjharte)i</td>
<td>–</td>
<td>–</td>
<td>0.871</td>
</tr>
<tr>
<td>junxuan20170516 (chenxj1005)i</td>
<td>–</td>
<td>3D</td>
<td>0.865</td>
</tr>
<tr>
<td>MEDICAi (bharadwaj)i</td>
<td>–</td>
<td>3D</td>
<td>0.862</td>
</tr>
<tr>
<td>Ethan20161221 (ethanhwang2012)i</td>
<td>–</td>
<td>–</td>
<td>0.856</td>
</tr>
<tr>
<td>resnet (QiDou) (Dou et al., 2017b)</td>
<td>NVIDIA Titan X × 1</td>
<td>3D</td>
<td>0.839</td>
</tr>
<tr>
<td>CCELargeCubeCnn (Intel,wuhui)online</td>
<td>CPU</td>
<td>–</td>
<td>0.833</td>
</tr>
<tr>
<td>Ours</td>
<td>NVIDIA 980 Ti × 1</td>
<td>2D</td>
<td><strong>0.880</strong></td>
</tr>
</tbody>
</table>

Fig. 5. Number of detected nodule candidates before and after candidate merging process with different cut-off thresholds of prediction probability for all 888 LUNA16 scans.

FPs relationship. A predicted nodule candidate will be counted as a true positive (TP) if it stays within five voxels to the real nodule center. The competition performance metric (CPM) score (Niemeijer et al., 2011) is calculated by taking the average detection sensitivity of seven predefined FPs per scan: 0.125, 0.25, 0.5, 1, 2, 4, and 8.

With regards to the nodule segmentation, we evaluate the segmentation similarity between automated segmentation and groundtruth mask through the calculation of the dice coefficients (Sørensen, 1948), which is defined by the following equation.

\[
DSC = \frac{2TP}{2TP + FP + FN}
\]

3.3. Nodule detection

After nodule candidate detection by Faster R-CNN, we apply a merging step to combine the candidates that are geometrically close to each other. The number of candidates before and after the merging process is shown in Fig. 5. The number of candidates decreases by 70.9%, 70.9%, 70.7%, and 70.3% at cut-off prediction probability threshold of 0.2, 0.4, 0.6, and 0.8, respectively, which verifies the superiority of using the merging operation as a simple postprocessing step in 2D nodule candidate detection. The FROC curve of our Faster R-CNN based nodule candidate detection is shown in Fig. 6a. When the threshold of prediction probability is set as 0.5, we achieve the sensitivity of 95.2% with an average of 19.8 FPs per scan, which outperforms the traditional nodule detection algorithm in LUNA16 with an overall sensitivity of 94.4% with an average of 620.6 FPs per scan. Our proposed merging method massively reduces the burden of FP reduction task. The FROC curves after 2D CNN based FP reduction are shown in Fig. 6b–d by taking initial candidate FP level of 10 (L10), 15 (L15), 20 (L20) from Faster R-CNN results. We list the sensitivity at each initial candidate FP level using our nodule detection method in Table 1 with CPM score of 0.866, 0.875, and 0.880 for L10, L15, and L20, respectively. The best CPM is achieved in L20 with sensitivity of 91.4% and 94.6% at an average of 1 and 4 FPs per scan, respectively. The number of trainable parameters of Faster R-CNN, FP reduction, and modified FCN networks are 2.4 × 10^6, 7.1 × 10^4, and 1.3 × 10^4. The 3D U-net-based architecture in CASED (Jensen et al., 2017), which has identical CPM value compared to our approach, has approximated 6.6 × 10^4 learning parameters. Our Faster R-CNN + CNN model for detection has about 36% parameters of CASED model. By investigating different initial candidate FP levels from detecting network, these results show that a less conservative initial candidate selection threshold yields stronger performance (better CPM value), when FP reduction is employed. Moreover, a more progressive initial candidate selection threshold only performs better sensitivity at smaller FP levels (less than 0.25 FPs per scan in our case) in FROC curve.

Comparing with the top 15 models under the nodule detection track in LUNA16 challenge that employed 3D CNNs, our 2D nodule detection method successfully outperforms some of the 3D methods (Row 10 to 15 in Table 1) while keeping the computation and memory usage efficiency. Different from the methods (Row 1, 2, 3, and 5) using multiple powerful GPUs, our 2D model shows the capability of using cost-efficient GPUs. For instance, compared with Zhu et al., 2018 that employed a 3D Faster R-CNN network with 3D dual-path blocks and U-net-like encoder-decoder structure to compactly and effectively exploit features, our light and straightforward approach has much smaller number of learning parameters, thus the whole 2D slices that contain the entire natural contextual information can be fed into the training without considering the GPU memory limitation. They tried to save memory consumption by cropping 3D patches with pixel size 96 × 96 × 96 from original scans, which may also add computation burden at the testing phase.

Nodules that are successfully detected by candidate detection network but eliminated by FP reduction network are shown in Fig. 7. About 4% of correctly detected nodules in our testing set are falsely removed. These nodules present either irregular shapes or ambiguous boundaries. Bringing more data targeting these nodule representatives might potentially boost the sensitivity performance.

3.4. Nodule segmentation

Not considering FPs, the mean and standard deviation of dice coefficients (DSC) regarding the nodule segmentation using FCN2s,
FCN4s, and FCN8s are presented in Table 2, which quantitatively proves that FCN2s with lower level convolution will potentially yield higher resolution and preserve the boundary information of the nodule. Therefore, higher DSC is obtained with the comparison to groundtruth mask. This proves our assumption that using lower level convolutional layers can help improve the segmentation accuracy. The higher level deconvolutional results provide the rough nodule locations and shape, while the lower level deconvolutional results provide the nodule boundary details.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>The mean and standard deviation of segmentation dice coefficients among FCN2s, FCN4s, and FCN8s.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>FCN2s</td>
<td>0.793</td>
</tr>
<tr>
<td>FCN4s</td>
<td>0.723</td>
</tr>
<tr>
<td>FCN8s</td>
<td>0.674</td>
</tr>
</tbody>
</table>

Fig. 6. The FROC curve of our nodule detection results. (a) FROC curve of Faster R-CNN based nodule candidate detection without FP reduction. (b) FROC curve of FP reduction by taking initial candidate FP level of 10 in (a) as initial candidates marked by green square. (c) FROC curve of FP reduction by taking initial candidate FP level of 15 in (a) as initial candidates marked by blue square. (d) FROC curve of FP reduction by taking initial candidate FP level of 20 in (a) as initial candidates marked by purple square (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Fig. 7. Nodules in the testing set that are eliminated by FP reduction network. Red dashed circles mark the nodule locations. The top row (a to c) are nodules in subset 8, and the bottom row (d to h) are nodules in subset 9 (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).
Fig. 8. Visualization results of our proposed nodule segmentation system with various anatomical characteristics. The columns 1–3 marked by red rectangle represent three isolated nodules, the columns 4–6 marked by light blue rectangle denote one juxta-pleural (column 4) and two juxta-vascular (columns 5 and 6) nodules, and the columns 7 and 8 marked by light purple rectangle show one subsolid nodule with center excavation and one ground glass opacity (GGO) nodule. The first row represents original nodule patches after Faster R-CNN detection and FP reduction with predicted bounding boxes marked by solid red squares as well as the classification probabilities in light blue background. The second row represents the corresponding annotations by radiologists. The manual segmentations are emphasized by red masks. The third to fifth rows denote nodule segmentation results by FCN2s, FCN4s, and FCN8s, respectively. The fully-automated segmentations are emphasized by yellow masks. The white decimals implicit the dice coefficients for each segmentation compared to groundtruth markings (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 3
Performance of the proposed method and other state-of-the-art approaches.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Nodule amount</th>
<th>IoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tachibana and Kudo (2006)</td>
<td>23</td>
<td>50.7 ± 21.9%</td>
</tr>
<tr>
<td>Wang et al. (2009)</td>
<td>64</td>
<td>58%</td>
</tr>
<tr>
<td>Messay et al. (2015)</td>
<td>68</td>
<td>63 ± 16%</td>
</tr>
<tr>
<td>Kubota et al. (2011)</td>
<td>23</td>
<td>69 ± 18%</td>
</tr>
<tr>
<td>Tan et al. (2013)</td>
<td>82</td>
<td>59 ± 19%</td>
</tr>
<tr>
<td>Lassen et al. (2015)</td>
<td>23</td>
<td>65%</td>
</tr>
<tr>
<td>Messay et al. (2015)</td>
<td>19</td>
<td>52 ± 7%</td>
</tr>
<tr>
<td>Wang et al. (2017)</td>
<td>40</td>
<td>50 ± 14%</td>
</tr>
<tr>
<td>Ours</td>
<td>77</td>
<td>71.70 ± 19.89%</td>
</tr>
<tr>
<td></td>
<td>493</td>
<td>69.23 ± 33.82%</td>
</tr>
<tr>
<td></td>
<td>223</td>
<td>70.24 ± 12.04%</td>
</tr>
</tbody>
</table>

Table 3 shows the comparisons between our proposal segmentation result and other researchers’ results based on the LIDC dataset. Since these methods use intersection over union (IoU) as the measurement metric, we report our results based on the same measurement in this table. Among all the listed approaches, our method achieves high performance on a relatively large amount of testing dataset.

In clinical practice, nodules have various types of clinical characteristics. The capability of segmenting large variety of nodules is necessary. Nine nodule characteristics are given in LIDC groundtruth markings, reflecting nodules’ calcification, malignancy level and so on. We chose four representative characteristics and separate our testing set into different groups according to the corresponding characteristic scores. Based on the quantitative results in Table 4, our segmentation algorithm possesses the robustness of processing various types of nodules with similar performance.

Table 4
Dice coefficients on different nodule groups based on clinical characteristics. Note that nodules in the testing set are grouped based on their clinical characteristic scores. The numbers in square brackets represent the number of nodules in the corresponding group. We average the characteristic scores from four radiologists.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Scores</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcification</td>
<td>–</td>
<td>–</td>
<td>0.794</td>
<td>0.817</td>
<td>0.788</td>
<td>0.795</td>
<td></td>
</tr>
<tr>
<td>Sphericity</td>
<td>–</td>
<td>–</td>
<td>[10]</td>
<td>[4]</td>
<td>[1]</td>
<td>[208]</td>
<td></td>
</tr>
<tr>
<td>Spiculation</td>
<td>0.784</td>
<td>0.740</td>
<td>0.823</td>
<td>0.781</td>
<td>0.798</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[1]</td>
<td>[19]</td>
<td>[72]</td>
<td>[76]</td>
<td>[55]</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.783</td>
<td>0.824</td>
<td>0.779</td>
<td>0.874</td>
<td>0.858</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[157]</td>
<td>[36]</td>
<td>[14]</td>
<td>[6]</td>
<td>[10]</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Malignancy</td>
<td>0.737</td>
<td>0.780</td>
<td>0.807</td>
<td>0.812</td>
<td>0.800</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[14]</td>
<td>[62]</td>
<td>[71]</td>
<td>[43]</td>
<td>[33]</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

Visual comparisons between groundtruth, FCN2s, FCN4s, and FCN8s are shown in Fig. 8 by displaying several representative nodules. Our nodule segmentation algorithm can precisely segment large variety of nodules with different anatomical characteristics. Due to the complicated boundary pattern, the DSC for poor-circumscribed (juxta-pleural and juxta-vascular) and fuzzy-boundary (ground glass opacity) nodules is a little worse than well-circumscribed nodules. But since the main components of such nodules are successfully segmented, the quantitative performance decay will not impact the robustness in terms of visualization.

Because we are conducting an end-to-end nodule segmentation directly from raw CT scans, the combined dice coefficient including FPs can reflect the overall performance. The false positives (FPs) can be involved in both detection and segmentation steps. For the
segmentation algorithm, the FPs can occur when the algorithm falsely detects other structures (vessels, airway walls, etc.) while avoiding the true nodules, which brings the dice coefficient to zero on FPs (Type I). For the detection algorithm, the FPs are apparently added when the algorithm fails to separate FPs from detected candidates (Type II). Leveraging from our full CAD system (detection + segmentation), the Type I FPs are eliminated by detection network since the center coordinates far from detected candidates will be removed after the postprocessing step illustrated in Section 2.2.6. Therefore, only Type II FPs are remained. We can simply calculate the combined dice score by considering all positive detected regions (TPs and FPs). For the given testing set, we choose sensitivity of 91.4% with 1 FP per scan in the detection step and feed all the candidates in the segmentation network, the end-to-end dice score will be decreased to 0.426 (0.793 if we exclude the Type II FPs). Therefore, the combined dice coefficient provides the convention to be compared to other end-to-end nodule segmentation algorithms.

3.5. Execution performance

As for execution perspective, we demonstrate the efficiency of the proposed method by running our automated nodule segmentation system for all 888 LUNA16 scans. Fig. 9 shows statistical box plots regarding nodule detection (Faster R-CNN + CNN) and nodule segmentation (FCN5s). Note that the execution time is scan wise, so there are some variations due to the variations of slice numbers. The mean values and standard deviations are indicated in Table 5. Our end-to-end fully-automated nodule segmentation system achieves an average of 16 s per scan without any human interventions.

4. Conclusions

In this paper, we developed a four-step pure CNNs based pulmonary CAD algorithm on thoracic CT scans, which can automatically and efficiently segment lung nodules with reasonable amount of FPs. Even though there is still a performance gap compared to some state-of-the-art methods that involved much denser 3D convolutional layers with much more complex designs, our results and evaluations demonstrated the capability of using pure 2D CNNs on a standard workstation to detect and segment pulmonary nodules with high performance. Such automation and efficiency significantly facilitate the translation from scientific researches to real applications on the computerized lung nodule segmentation trend. For the future outlook, further evaluation based on independent testing dataset is desired to perform a more comprehensive comparison. Besides, for any 2D detection scheme implemented on volumetric imaging, the model performance is also critical when being applied across an entire volume. We will assess it in the future. Finally, it is attractive to implement more advanced 2D CNNs architectures (such as generative adversarial network in Goodfellow et al., 2014) to further boost the performance.

Conflict of interest statement

No conflict of interest.

Acknowledgement

This work was supported by US National Science Foundation (award number 1429708 and 1827745).

References


Table 5

<table>
<thead>
<tr>
<th></th>
<th>Mean (secs)</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection (Faster R-CNN + CNN)</td>
<td>15.319</td>
<td>8.162</td>
</tr>
<tr>
<td>Segmentation (FCN5s)</td>
<td>0.917</td>
<td>0.835</td>
</tr>
</tbody>
</table>


