Journal of Ambient Intelligence and Humanized Computing (2023) 14:14817–14827
https://doi.org/10.1007/s12652-018-01170-5

ORIGINAL RESEARCH

Supervoxel based weakly-supervised multi-level 3D CNNs for lung nodule detection and segmentation

Yuanli Feng[1,2] · Pengyi Hao[2,3] · Peng Zhang[4,5] · Xinguo Liu[1,2] · Fuli Wu[2,3] · Hongwei Wang[4]

Received: 18 January 2018 / Accepted: 29 December 2018 / Published online: 6 March 2019

Abstract
Pulmonary nodule detection and segmentation are two important works for early diagnosis and treatment of lung cancer. The work of detection is to locate pulmonary nodules in a given chest CT scan, and the segmentation aims at extracting all the voxels from a CT scan within each nodule's space. This paper propose a novel framework to process both nodule detection and segmentation integrately, which is implemented as the combination of SLIC supervoxel segmentation and CNN classification. The learning of CNN just require weakly labeled data annotations, where only a single coordinate is provided for each annotated nodule as the ground truth. The CNN architecture is designed as a 3D multi-level framework, which is able to comprehensively recognize nodules with variant sizes and shapes. Experiments on the dataset of LUNA16 challenge expressed prominent detecting performance, demonstrating the necessity and efficiency of 3D CNN architecture and multi-level framework for computer-aided detection of pulmonary nodules. Meanwhile the evaluation of segmentation presented impressive performance, producing elegant shapes of real nodules, which proves the great efficiency of SLIC technique.

Keywords Pulmonary nodule detection · Pulmonary nodule segmentation · SLIC · 3D convolutional neural networks

1 Introduction

Lung cancer is a kind of severely mortal illness worldwide, lying at the top for both mortality and morbidity among all the cancerous diseases. In order to help lung cancer prediction in the early stage, computer-aided diagnosis (CAD) systems are often used to assist radiologists to read medical images more precisely and efficiently. The works of CADs include various medical image processings such as denoising (Mingliang et al. 2016), segmentation (Ronneberger et al. 2015; Tang et al. 2016), and detection or characterization of medical lesions (Tajbakhsh et al. 2016; Fakoor et al. 2013; Sirinukunwattana et al. 2016). CADs for lung cancer always focus on lung nodule analysis from volumetric thoracic computed tomography (CT), where lung nodule is a kind of granuloma regarded as an important reference factor in lung cancer diagnosis.

Nodule detection and segmentation are two main processes in lung nodule analysis, where detection is to find out and localize nodules in CT scans by calculating a set of estimated coordinates to help release the manual cost of radiological screening, and segmentation is to give a precise voxel-wise presentation of nodule shape to help pathological diagnosis. Their main details are as follows.

1.1 Nodule detection

An automated pulmonary nodule detecting system usually consists of two steps: (1) candidate detection and (2) false positive reduction. The process of candidate detection aims to get a set of nodule candidate locations with the highest reachable sensitivity, which results in a large number of false positives. Then during the stage of false positive reduction,

✉ Peng Zhang
zhangpeng1121@aliyun.com

Yuanli Feng
ylfeng@zju.edu.cn

1 State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058, China

2 Real Doctor AI Research Centre, Zhejiang University, Hangzhou 310058, China

3 School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

4 School of Economics and Management, Tongji University, Shanghai 200092, China

5 Tongji University Affiliated to Shanghai Pulmonary Hospital Thoracic Surgery, Shanghai 200092, China

candidates are filtered by a classifying scheme, where the numerous false positives are widely eliminated, leaving a precise detection result of true nodules.

The common candidate detecting methods are often simple and effective. Curvature thresholding and centering (Murphy et al. 2009) are proceeded to extract nodules with sphere-like shape, and some other methods just use double thresholding and morphological operations (Jacobs et al. 2014; Setio et al. 2015) to find out other oddly shaped nodules. In the step of false positive reduction, a set of delicate features are extracted, and one or more supervised classifiers are built for them to distinguish more detailed informations between true nodules and non-nodule candidates (Murphy et al. 2009; Jacobs et al. 2014; Messay et al. 2010).

Although above methods can help improve the reading efficiency of radiologists, they are far from practical applications. A significant reason is that nodules vary from a wide range in shape, size, and texture, while existing simple classifiers cannot entirely recognize such kind of complex features precisely. Another reason is the existence of nodule-like false positives, where some general tissues within lung space look too similar to true nodules. Therefore, it's hard to both eliminate the false positive rate and maintain a high sensitivity simultaneously.

Recently, the newly popularized deep learning technique has brought such applications into a new stage of research. Convolutional neural networks (CNN) is a kind of image processing tool with sets of arithmetic units concatenated layer-by-layer to analyze inputs with various kinds of features, which can easily adapt to applications of medical image analysis, improving accuracies of nodule detecting performance. Recent works have applied CNNs to precise classification of nodule candidates. Yang et al. (2016) used a 4-layer 2D CNN with input of 50×50 cropped local patches to proceed classification in LIDC-IDRI dataset. Setio et al. (2016) proposed a multi-view approach by cropping multiple patches along nine symmetric planes of the local volume centering at each given candidate location, and put them separately into a 5-layer 2D CNN with fusion framework appended in the tail. Ramaswamy and Truong (2017) implemented nodule classification with pretrained AlexNet and GoogleNet, and extended AlexNet into 3D architectures to analysis CT scans with the original dimension.

Early researches of CADs based on deep learning have often used 2D CNNs for their low cost of time and memory, and some pre-trained networks can improve the efficiency of training process. However, methods based on 2D CNNs still could not take full advantage of those with 3D CNNs to recognize nodules in complicated environments with variant characteristics, where 3D CNNs can encode richer spatial information and extract more representative features via training on complete 3D CT samples than those 2D models. Up to now, 3D CNN is still in an early stage of medical applications, and only a few 3D variants of CNNs have been lately proposed for pulmonary nodule analysis (Dou et al. 2017). Because of the high memory and time cost of 3D CNNs, and the lack of professionally labeled training data, the performance of 3D CNNs are still limited, and need to be further explored.

1.2 Nodule segmentation

In analysis of pulmonary nodules, the shape of nodule is an important factor for diagnosis of nodule's malignancy, where a segmentation process is needed. Traditional robust nodule segmentation method was proposed by Kuhnigk et al. (2006), with the combination of a number of morphological operations on the cubic volume of interest to acquire a robust nodule segmentation. Such method is applied after candidate detection, to provide information of shapes for further false positive reduction.

Recently, nodule segmentation methods based on deep learning have appeared. The segmentation with CNNs is hard to implement, for the lack of voxel-wise ground-truth labels, thus recent researches often emphasis on weakly-supervised applications. For example, Anirudh et al. (2016) preprocessed weakly labeled data by supervoxel clustering to obtain estimated segmentation labels, and trained a 3D CNN to classify voxels within given CT scans. Feng et al. (2017) used slice-level labeled samples to train a binary classifying network, and generated the segmentation result from filter weights of the last fully-connected layer as nodule activation maps (NAMs). Up to now, the accuracy of weakly-supervised segmentation approaches are still far from enough to help subsequent diagnosis, and is therefore in desperate need of further improvement.

In this paper, we propose a novel approach to conduct both nodule detection and nodule segmentation processes together. Instead of traditional candidate methods focusing on thresholding of density and curvature, we utilize an efficient supervoxel method named simple linear iterative clustering (SLIC) (Achanta et al. 2012) to produce candidate supervoxel clusters of pulmonary tissues, which is more accurate to separate nodular voxels from normal tissues. Then these supervoxels are classified with CNNs to denote whether they are within nodules or not. 3D CNN is used rather than 2D CNN to take full advantage of the 3D spatial information, and the multi-level framework is designed to recognize nodules with large variations of sizes, where the technique of multi-level is widely applied to deep neural networks for common image processing (Shao et al. 2014; Wu et al. 2016) to achieve better precision than the use of single deep networks, and can also perform well in medical imaging tasks such as pulmonary nodule detection. After classification, the class predictions of the supervoxels are combined and finally voxel-wise segmentation results of

nodules are produced, while the centers of nodules are easy to be calculated as the detection result. The 3D CNNs are trained on weakly-labeled lung nodule datasets with only a centering coordinate of each nodule annotated, or sometimes an approximate diameter provided, while the CNNs cooperating with SLIC can successfully proceed a voxel-level segmentation for each nodule.

Our main contributions are as follows: (1) We propose an integrate framework to implement both works of pulmonary nodule detection and segmentation in a single process; (2) We design 3 3D CNNs with different scales of receptive fields, and evaluate their performance with 2 different fusion methods to more extensively analyse the contextual information of the input; (3) We unite the SLIC and CNN classification process, to implement voxel-level nodule segmentation with only weakly-labeled training data, such as dataset provided by LUNA16 challenge, to release the drawback of the poorly provided fully-labeled data.

The rest of this paper is organized as follows. The main details of our method are described in Sect. 2. The experimental details and results are introduced in Sect. 3. Section 4 organizes the discussion, and the conclusions are finally drawn in Sect. 5.

2 Method

The main process of our method is illustrated in Fig. 1. It can be summarised as four steps: (1) CT image preprocessing, (2) supervoxel generation, (3) deep learning based supervoxel classification and (4) predicted supervoxel combination. The supervoxel generation can be regarded as a candidate detection process to produce a set of candidate coordinates for further classification, and supervoxel classification corresponds to the step of false positive reduction. Their main details are as follows.

2.1 Preprocessing

The input of our CAD system is a 3D CT image of a human's chest. Such kind of CT images are produced from variant medical facilities, so their format differ very much, which makes it hard to analyse their spatial and contextual features. Therefore, preprocessing is an essential step to fit large amounts of variant data into a conform standard.

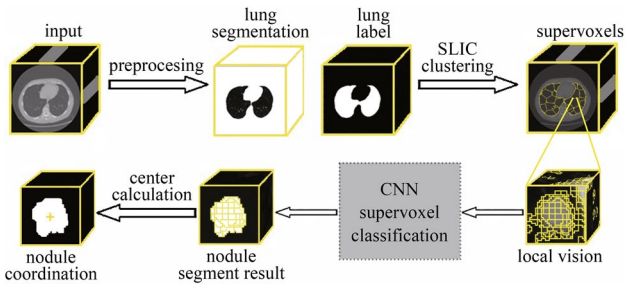

Fig. 1 The main process of our method

Preprocessing is always conducted by researchers for years, and there're some particular kind of implementations for this process. Firstly, CT values in CT scans should be converted into Hounsfield unit (HU) values sometimes, where HU is a common standard of CT image processing. For our data is of off-the-shelf format, this step is not necessary in this paper. Secondly, resampling is needed for CT scans with different resolutions, which proceeds mathematical interpolation to modulate them into the same per-voxel spacing. Thirdly, lung space segmentation is performed by connectivity analysis to extract voxels inside the lung, which eliminates unnecessary subsequent calculations on lung walls and the outside air. Finally, normalization and centering should always be applied for deep learning methods, because deep neural networks are fragile for complicated training gradients, while such two processes can limit the intensity of features and keep training gradients in control.

2.2 Supervoxel generation

In contrast to traditional density (Jacobs et al. 2014; Setio et al. 2015) and curvature (Murphy et al. 2009) thresholding techniques for candidate detection, we use SLIC (Achanta et al. 2012) to produce candidate locations.

SLIC is an adaption of k-means superpixel method, while two improvements are applied: (1) The search space for distance calculation is set proportional to the size of the source superpixel instead of the overall pixel region to efficiently reduce the computational cost; (2) The distance measure combines weighted color and spatial proximity to simultaneously control the sizes and compactness of the superpixels.

Original SLIC is applied to 2D natural images with 3 channels as RGB by range 0–255 per pixel, while pulmonary CT images are all 3D scans with HU value −1000 to 1000 per voxel. Fortunately, SLIC is a dimension-invariant method which can be easily extended to 3D supervoxels, where we need only alter the distance measurement to fit it into processes of CT scans.

For a CT scan with number of voxels N and number of target supervoxels k, the supervoxel grid length is set $S = \sqrt{N/k}$. Each voxel contains 2 types of information, which are HU value l and position $[x\,y\,z]^T$, then the distance between two voxels i and j is written as

$$\begin{aligned} d_s &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2} \\ d_c &= |l_i - l_j|_1 \\ D &= \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2 m^2} \end{aligned} \tag{1}$$

where S and m are defined as the maximum spatial and HU value distance within a given supervoxel. In practical, m is often controlled by users to weigh the relative importance between voxel value similarity and spatial proximity.

SLIC separates a CT image into a vast number of supervoxels, examples of local patches are shown in Fig. 2. Voxels within the same supervoxel are distributed in a dense region, and contain similar HU values. In ideal condition, voxels belonging to a nodule should be always from different supervoxels with voxels of non-nodule tissues, while voxels of the same nodule may also sometimes scatter into different supervoxels. To avoid rough clustering of voxels from small nodules, we carefully set the number of supervoxels to ensure the sizes of supervoxels lie in about $3 \times 3 \times 3$ corresponding to the smallest valuable nodules with diameter 3 voxels. We clipped the supervoxels by eliminating voxels with HU values below -600, for they represent empty areas without need of further analysis. Then each remaining supervoxel here is regarded as a nodule candidate, and is intended to be classified as whether "within a nodule" or not in Sect. 2.3.

SLIC is more efficient than sliding window technique, for it can give us a spatial clustered set of candidate positions instead of sliding positions, which do not counter any relational informations between voxels within the same nodule. This method is also sensitivity-preserving in contrast to traditional candidate detection methods (Jacobs et al. 2014; Setio et al. 2015; Murphy et al. 2009). Nevertheless, the false positive rate of such detection result often explode, which need to be solved in the next step of supervoxel classification.

2.3 Supervoxel classification: local volume extraction

In Sect. 2.2 we obtain a series of tissue clusters as supervoxels from a given CT scan, and here we need to accurately classify each supervoxel into two categories, that is whether it is within the region of a nodule or not. After this process, the false positive supervoxels are meant to be all classified as non-nodule and eliminated from detection results.

A simple false positive reducing process by 3D CNN is straightforward. For each supervoxel, we firstly extract a cubic local volume centering at the supervoxel's center, which contains a part of contextual information surrounding it, then put such volume into a designed 3D CNN to produce a binary classifying result for the centering supervoxel, and finally, combine the voxels of the true-nodule supervoxels to obtain the set of voxels within estimated nodules.

However, such simple process can not accurately recognize all the nodular supervoxels. A significant reason is that nodules' sizes and shapes vary broadly, which directly caused context inconsistency. In this case, an input region with a smaller nodule often contains more contextual information, and vice versa. If an input region contains too much redundant context, the classification model will be misled to pay more attention on voxels outside the nodule. Conversely if the region is too small for a large nodule, the surrounding environment of the nodule will not be sufficiently preserved, and the information within the nodule is even possible to be partially cut out.

In order to solve such kind of problems, Dou et al. (2017) proposed a multi-level anisotropic 3D CNNs model. In that method they proceeded a statistical analysis to nodule sizes, by which they designed 3 CNNs with corresponding 3 scales of receptive fields: $20 \times 20 \times 6$, $30 \times 30 \times 10$ and $40 \times 40 \times 26$ voxels (each included nodule sizes by 58%, 85% and 99% of the overall nodules from LUNA16 dataset). Their model produced good performance on the dataset of LUNA16 challenge and achieved a high rank there.

Nevertheless, a drawback for that approach is that the 3 dimensions of the data they processed are holding anisotropic spacings between each neighboring voxels, which results in a deformation to the real shape of nodules. Though their model can adapt to the common deformations on LUNA16 dataset, it can not fit other datas with different spacing proportions. To attain spacing robustness, we propose a multi-level isotropic model, where the scales of receptive fields for the 3 CNNs are respectively: $20 \times 20 \times 20$, $30 \times 30 \times 30$ and $40 \times 40 \times 40$ voxels.

2.4 Supervoxel classification: multi-level framework

The main structure of our multi-level 3D CNNs framework are shown in Fig. 3, where each network has four

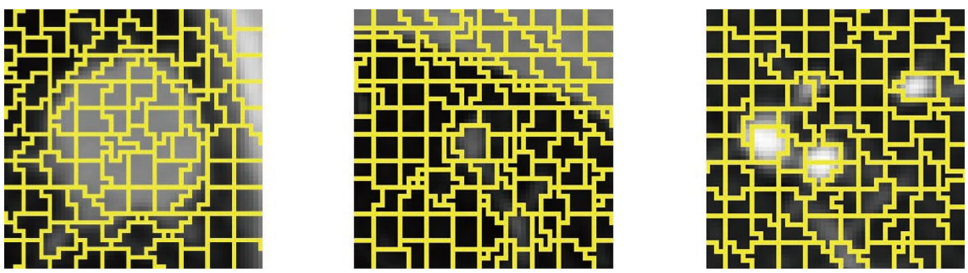

Fig. 2 SLIC segment on CT images, from left to right are respectively large nodule, small nodule and non-nodule patches

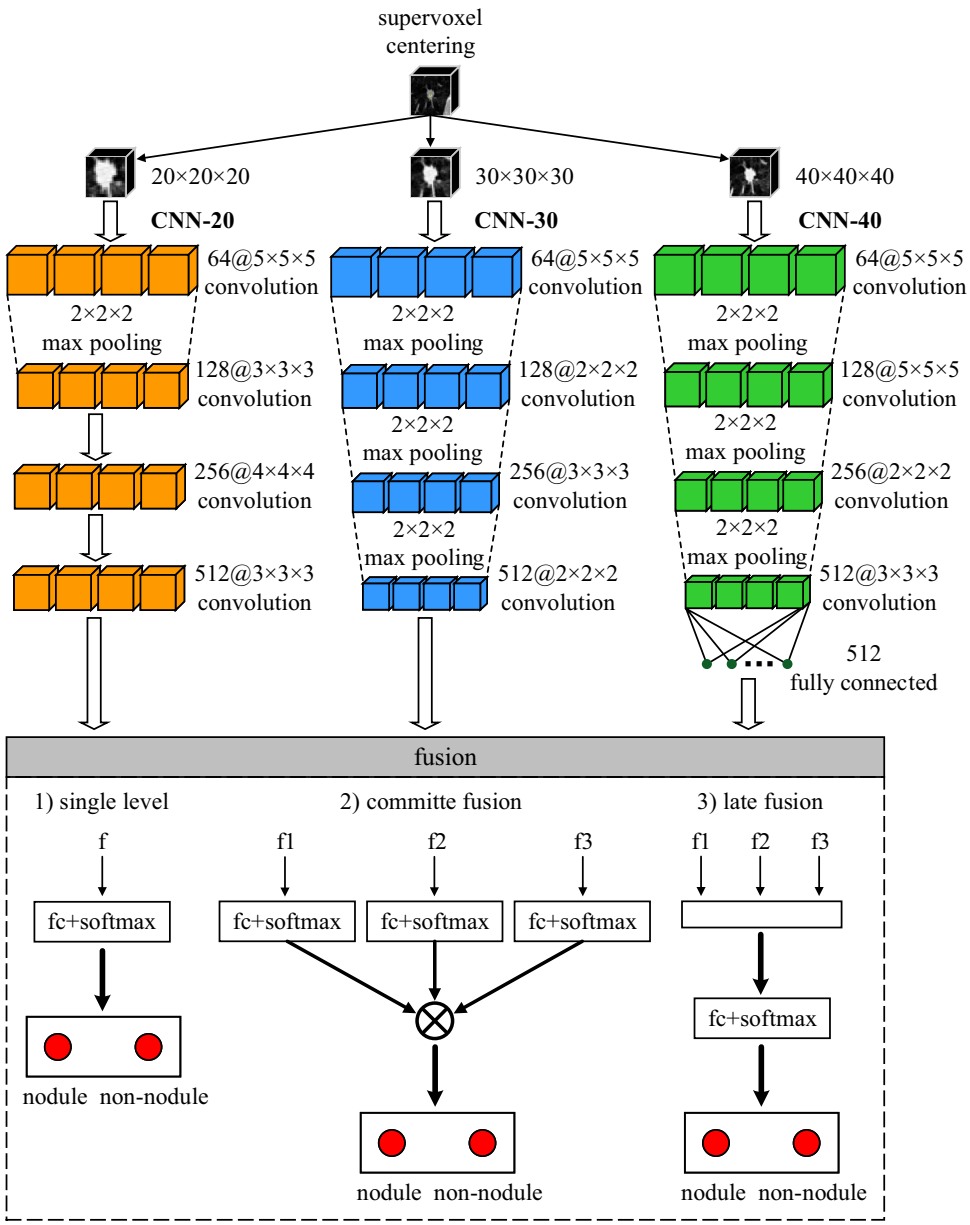

Fig. 3 The supervoxel classification framework. 3 CNNs with different scales of receptive fields are proposed to analyse a given supervoxel in cooperation. All the 3 networks have four convolutional layers, each layer with M kernels of size $N \times N \times N$ is denoted as $M@N \times N \times N$. After the last hidden layer of each network, there's a fully connected layer appended by a softmax layer to finally produce a probability of nodule or non-nodule classification, and a fusion technique is proceeded to combine their features and produce an integrate classifying result

convolutional layers. Both CNN-20 and CNN-30 contain one fully-connected layer, while the CNN-40 with larger receptive field contains 2 fully-connected layers. Batch normalization (Ioffe and Szegedy 2015) layers are inserted after each hidden layer to ensure a higher learning rate and reduce overfitting, with dropout layers (Hinton et al. 2012) appended to further reduce the performance of overfitting.

Each of the 3 architectures outputs a 2-D classifying prediction to nodule or non-nodule from the last softmax layer and a 512-D feature vector from the last hidden layer, then their outputs should be combined into a single classifying result for the original given supervoxel. We used two data fusion techniques, namely committee-fusion (Van Ginneken et al. 2015) and late-fusion (Prasoon et al. 2013). In committee-fusion, the predictions output by each softmax layer of the 3 architectures are combined into a single probability with weighed averaging or a specific product rule (Van Ginneken et al. 2015). While for late-fusion, the 3 features from the last hidden layer of the CNNs are concatenated into an integral feature vector and sent to a fully-connected layer to produce the result of probability.

2.5 Prediction combination

After supervoxel classification, the supervoxels within nodules are meant to be entirely found out, then their corresponding voxels are combined into estimated nodules with 18-connected seed dispersal, and therefore produce

the nodule segmentation result. For detection task, accurate coordinates within CT scans should be generated, so we calculated the center of the estimated nodules by averaging their inner voxel coordinates. Both detection and segmentation are proceeded by an integrate framework, with simply supervoxel techniques together with classifiers to implement works of other state-of-the-art approaches with separate complicated frameworks.

2.6 Training

The CNNs need training process to learn features of inputs with various contextual information. We trained the 3 CNNs with a combined dataset of LUNA16 and another vast patient series provided by Shanghai Pulmonary Hospital (SPH) in China. The nodules' annotations are used to make positive training samples. For LUNA16 provided a set of candidate locations for every CT scan, we utilized them to produce our negative training samples. For each nodule or non-nodule location, we extracted 3 cubic volumes centering at the given coordinate, with scales respectively corresponding to the receptive fields of the 3 architectures, then these labeled local volumes are sent to CNNs for training.

A significant problem during training is the class imbalance, where the number of negative samples is hundreds of times more than that of positive samples. We dealt with this in two ways, which are respectively data augmentation and loss adjustment.

2.6.1 Data augmentation

This process is often applied to release data imbalance and expand receptive domain of the models trained. In our experiments, we rotated the positive input volume by 90°, 180° and 270° around each of the 3 spatial axis, and appended flipping respectively along these 3 directions, which produced 13 training volumes for every positive sample.

2.6.2 Loss adjustment

The loss most commonly used for nodule detection is the cross entropy loss. Nevertheless, such form of loss is easy to be affected by data imbalance, because samples in the prominent class often provide much more importance on back-propagated gradients, which misleads networks to recognize non-nodule supervoxels better, resulting a high false positive rate. Therefore, Lin et al. (2017) proposed focal loss to help weak class gain emphasis in training gradients. The focal loss is an adaption of cross entropy, with a modulating factor $(1-p_t)^{\gamma}$ inserted to formulate the loss as:

$$FL(p_t) = -(1-p_t)^{\gamma} log(p_t) \tag{2}$$

where γ is a focusing parameter with $\gamma \geq 0$, and p_t is the correction factor. We set $\gamma = 2$ as recommended in Lin et al. (2017) for our experiments. For the ground truth label $y = \{0, 1\}$ and probability p produced by the classifier on the class with label $y = 1$, the correctness is defined as:

$$p_t = \begin{cases} p, & \text{if } y = 1; \\ 1-p, & \text{otherwise.} \end{cases} \tag{3}$$

By this form of loss, the importance of misclassified examples are significantly increased in positive samples, while the effect of the majority redundant negative samples are greatly reduced for more balanced gradients.

3 Experiments

3.1 Preprocessing and supervoxel generation

In preprocessing, we resampled all the input CT scans into an uniform scale by 1 mm spacing between each neighboring voxels. For local volumes input into networks, we performed normalization with upper bound 512 and lower bound −1024, and centering with mean value 0.25. Normalization was also conducted before the generation of supervoxels. For testing process, we segmented the lung space in tested CT scans by connectivity analyse, with morphological closing appended to prevent omission of nodules located on lung walls.

Then SLIC was proceeded to over-segment voxels within lung space. The preprocessing before SLIC is a normalization to make the HU compactness among voxels easy to control, where we set the compacting factor to $m = 0.001$ for the following SLIC process. After clustering, we appended a postprocessing step as thresholding on the range of original HU values for −600. Such threshold is consulted from pulmonary specialists, while only voxels with HU above it is possible to locate within nodules. Thus we filtered out voxels below this threshold, simultaneously eliminating a large number of supervoxels and reduced the computational cost. The finally remaining supervoxels were all of inner tissues, and were to be further filtered in the next process of false positive reduction.

3.2 Datasets

To train the classifying CNNs, we used two annotated pulmonary datasets. One is the LUNA16 dataset, and the other is the provided SPH dataset.

The dataset released by LUNA16 challenge held in conjunction with ISBI 2016 is adapted from LIDC-IDRI dataset, containing 888 CT scans annotated by 4 radiological experts voxelwise. In LUNA16 nodules with diameter less than 3

mm or voted by less than 3 radiologists are eliminated, and each annotation of a nodule is simplified to two informations as a centering coordinate and the corresponding diameter in mm units.

For datas of SPH, the CT scans are circulated among nosocomial departments in a complicated procedure while we've got just 892 CT scans at present, and a pulmonary expert is engaged in annotations by presenting an approximate centering coordinate inside each nodule's area. For there are no other characteristics like nodule diameters provided from this dataset, it is simply used during training, and evaluations are conducted on only the dataset of LUNA16.

3.3 Evaluations

Our multi-level framework consists of 3 networks with different structures, and two kinds of fusion techniques. We conducted evaluations on each of these networks, and on the combined structure with each of the fusion methods.

Our experiments totally contain two tasks, which are pulmonary nodule detection and segmentation. The overall process is a streamline, while the detection result of a nodule is calculated at last as the center of the previous segmented nodules.

Figure 4 presents examples of segmentation. We filtered the predicted supervoxels with the threshold of 0.2, which eliminated predictions outside nodules. The remaining supervoxels were combined into a nodule segmentation mask, where the supervoxels close to a nodule's center are of higher confidence, and supervoxels near the border are predicted lower. The results show that our framework can obtain accurate shape of nodules with different sizes and textures, even difficult nodules attaching to lung walls can be differentiated well. In the meantime, some ambient tissues suspected to lie within nodules were successfully recognized, but of a relatively low confidence to support further analysis.

Examples of successfully detected nodules are listed in Fig. 5. It can be observed that our framework is able to recognize nodules with variant sizes, shapes and locations in a pretty high confidence. For quantitative evaluation, the predicted locations were compared with ground-truth locations provided from annotations, where an estimated location was considered as true positive if it is within the radius of a true nodule's center. The competition performance metric (CPM) (Niemeijer et al. 2011) was calculated as the average sensitivity at seven corresponding false positive rates as respectively 1/8, 1/4, 1/2, 1, 2, 4 and 8 FPs per scan, and free receiver operation characteristic (FROC) analysis was performed by setting thresholds on the rawly predicted probabilities according to the target FPs per scan, and obtaining the statistical sensitivities correspondingly.

The FROC curves of our different modules are shown in Fig. 6, where the network of CNN-30 generates the best scores among the 3 single models, and the committee-fusion of 3 networks exceeds the single CNN-30 to reach the top performance. Table 1 lists the sensitivities of our different modules at different false positive rates, both of the modules CNN-30 and committee-fusion achieve sensitivities beyond 90% under the false positive rate of 4 and 8 per scan, and the sensitivities of committee-fusion is generally higher than CNN-30 at most of the false positive rates, which proves the effectiveness of cooperation among networks with different receptive fields.

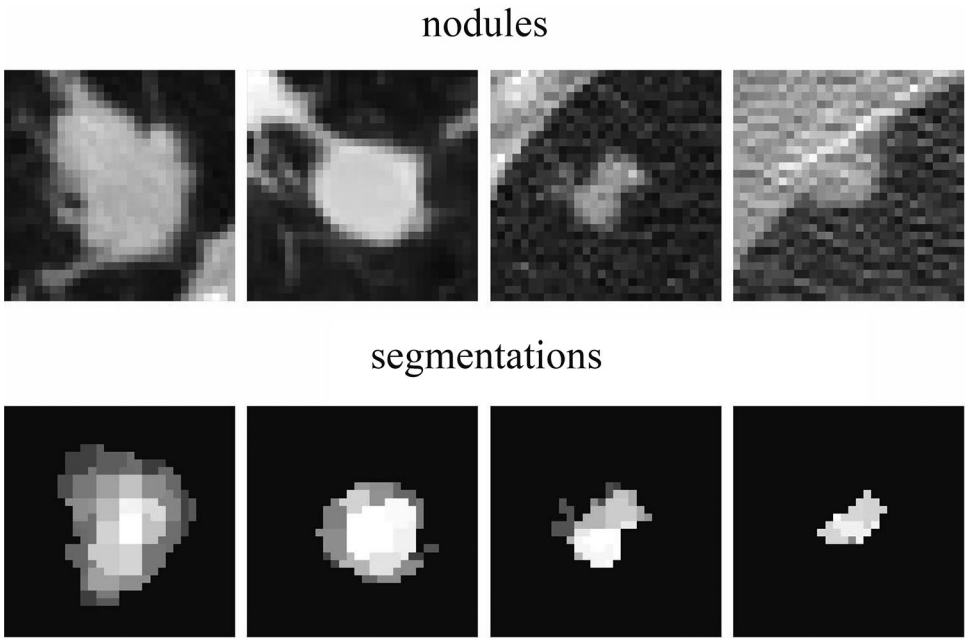

Fig. 4 Examples of pulmonary nodule segmentation results of our framework. The top row lists patches as representative transverse planes of nodules with variant sizes and shapes, and the bottom row presents their segmentation masks. The lighter mask indicate a higher confidence of nodules' supervoxels

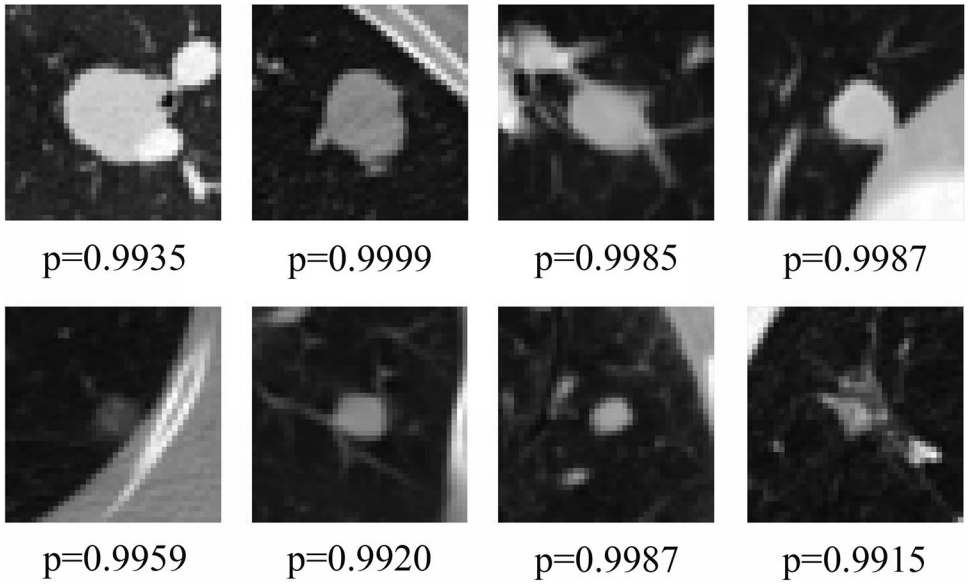

Fig. 5 Examples of pulmonary nodule detection results of our framework. Each patch is a transverse plane of an annotated nodule, and the p value presented below is its estimated probability to be the nodule

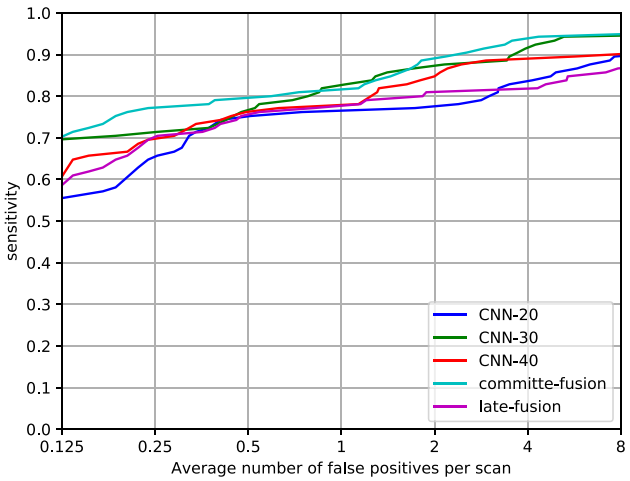

Fig. 6 FROC curves of different CNN architectures and fusion methods

Table 1 Sensitivities of our architectures respectively under different false positive rates

FP/scan	CNN-20	CNN-30	CNN-40	Committee	Late
0.125	0.552	0.693	0.581	0.695	0.571
0.25	0.648	0.705	0.695	0.771	0.695
0.5	0.743	0.762	0.762	0.790	0.752
1	0.762	0.819	0.771	0.810	0.771
2	0.771	0.867	0.848	0.886	0.809
4	0.829	0.914	0.886	0.933	0.810
8	0.894	0.942	0.896	0.943	0.867

Comparisons of sensitivities and CPM scores among different approaches are listed in Table 2. Our method achieves the highest CPM score among these methods, exceeding previous multi-level method. Our sensitivity at false positive rate of 8 per scan reaches 0.943 beyond all the other methods. The sensitivity of our framework at low false positive rates also ranks top among other approaches, reaching 0.695 at 0.125 false positives per scan.

4 Discussion

We propose a novel approach based on SLIC (Achanta et al. 2012) and CNN classifier to proceed pulmonary nodule detection and segmentation. SLIC divides pulmonary tissues into a large number of small supervoxels, and CNN classifier discriminates these supervoxels as whether lying within nodules or not. Then the classifying result forms a nodule segmentation mask, and finally the nodules' centers are decided as the center of each connected mask. The works of nodule detection and segmentation are conducted by an integrate process, instead of a separate implementation.

The CNN classifiers are only trained on weakly labeled datasets, with a single coordinate within each nodule or sometimes an approximate diameter provided additionally, while our framework with such kind of training data can produce a voxel-level nodule segmentation result.

For CNN architecture, we construct a 3D multi-level CNNs framework and apply two different fusion techniques to support comprehensive experiments. In contrast to previous 2D based CNNs, 3D CNNs are capable to analyse 3D spatial features around nodule space, and recognize nodules more precisely. The multi-level CNNs framework consists

Table 2 Results of the false positive reduction track in ISBI LUNA16 challenge

Method	CNN type	0.125	0.25	0.5	1	2	4	8	CPM score
DIAG (Setio et al. 2016)	2D	0.636	0.727	0.792	0.844	0.876	0.905	0.916	0.814
iitm03	2D	0.394	0.491	0.570	0.660	0.732	0.795	0.851	0.642
luna16cad	3D	0.640	0.698	0.750	0.804	0.847	0.874	0.897	0.787
LungNess	2D	0.453	0.535	0.591	0.635	0.696	0.741	0.797	0.635
UACNN	2D	0.655	0.745	0.807	0.849	0.880	0.907	0.925	0.824
NResNet (Dobrenkii et al. 2017)	3D	0.517	0.602	0.720	0.788	0.822	0.839	0.856	0.735
CUMedVis (Dou et al. 2017)	3D	0.677	0.737	0.815	0.848	0.879	0.907	0.922	0.827
Ours	3D	0.695	0.771	0.790	0.810	0.886	0.933	0.943	**0.833**

The bold value indicates the key expression of the performance of our approach, which present superiority to other approaches listed

of 3 CNNs with different scales of receptive fields to handle nodules with different sizes, and the fusion techniques conduct a more comprehensive recognition of variant nodules. In our experiments, the committee-fusion produced the best result, exceeding other methods in the dataset of LUNA16 challenge. However, the late-fusion technique did not perform prominently, which may result from an insufficient quantity of training samples, while the architecture of multi-level 3D CNNs concatenated by an integrate fully-connected layer requires more training data to reduce overfitting, thus further utilization of more annotated 3D samples can be possible to improve this technique.

Figure 7 lists some nodule detection results with low confidence. The left group shows some real nodules with relatively low predictions. while they often present irregular shapes or obscure boundaries, and a threshold of 0.6 can be employed to successfully retrieve these challenging cases. Nevertheless, some of the real nodules are also poorly predicted by our framework, as shown in the right group, where these kind of nodules often contain complex surrounding characteristics, or featureless textures with too little occupancy spaces, that an emphatic training can be proceeded in further experiments to increase the detecting performance on these outliers.

Our CNN classifier can also work for simple false positive reduction, where SLIC can be regarded as a form of candidate detecting process preserving full sensitivity but high false positive rate per scan. Therefore, the CNN architectures can be separated out from the whole framework independently and combined with any other candidate detecting process as another work of pulmonary nodule detection.

5 Conclusion

In this paper, a novel framework based on efficient clustering cooperating with 3D CNN classification is proposed to perform computer-aided detection and segmentation of

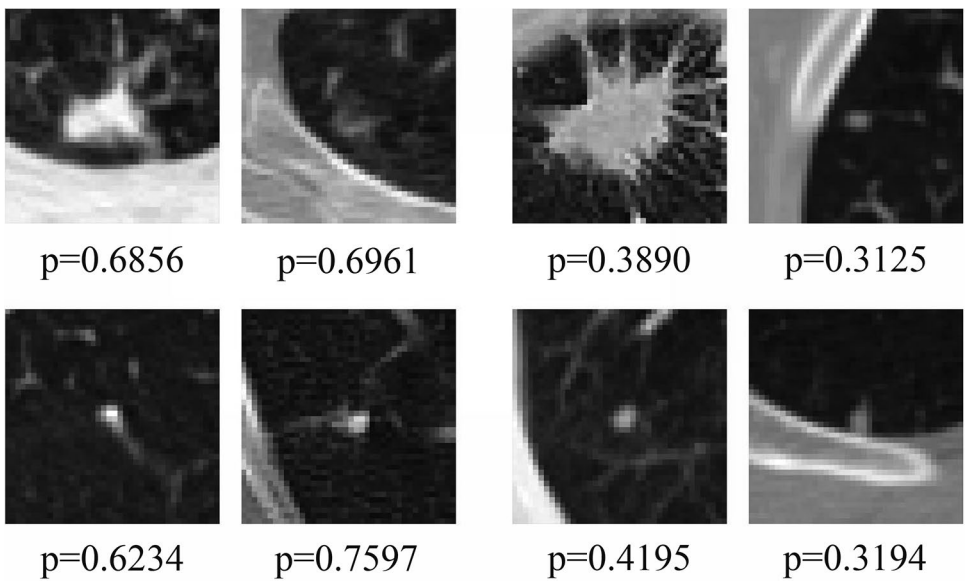

Fig. 7 Examples of pulmonary nodule detection results of our framework. Each patch is a transverse plane of an annotated nodule, and the p value below is its prediction as the confidence to be a nodule generated by our framework

pulmonary nodules from volumetric CT scans. The experiment is just supported by weakly labeled datas, while a voxel-level nodule segmentation can be performed by our framework. The CNN classifier is designed as a multi-level framework, consisting of 3 CNNs with different sizes of receptive fields to process nodules with variant sizes and shapes, and consequently our nodule detection system has conducted impressive performance in the dataset of LUNA16 challenge.

Future research still need to promote the accuracy of pulmonary nodule detection and segmentation, and the quantization of nodules' malignancy is also necessary for further investigation to perform a more comprehensive pathological diagnosis for patients. The future development requires active cooperation with pulmonary experts. With professional knowledge exchange, and vast dataflow containing valuable spatial and pathological information, experiments can be more efficiently proceeded in future works.

Acknowledgements This research is funded by the Zhejiang Provincial Natural Science Foundation of China under Grant no. LY18F020034, the Natural Science Foundation of China under Grant no. 61801428, no. 61872317. The work of this paper is also supported by Shanghai Pulmonary Hospital, who provided a vast annotated pulmonary dataset to help our experiments.

References

Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S (2012) Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell 34(11):2274–2282

Anirudh R, Thiagarajan JJ, Bremer T, Kim H (2016) Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data. In: Medical imaging 2016: computer-aided diagnosis, international society for optics and photonics, vol 9785, p 978532

Dobrenkii A, Kuleev R, Khan A, Rivera AR, Khattak AM (2017) Large residual multiple view 3D CNN for false positive reduction in pulmonary nodule detection. In: Computational intelligence in bioinformatics and computational biology (CIBCB), 2017 IEEE conference. IEEE, pp 1–6

Dou Q, Chen H, Yu L, Qin J, Heng PA (2017) Multilevel contextual 3-D CNNS for false positive reduction in pulmonary nodule detection. IEEE Trans Biomed Eng 64(7):1558–1567

Fakoor R, Ladhak F, Nazi A, Huber M (2013) Using deep learning to enhance cancer diagnosis and classification. In: Proceedings of the international conference on machine learning, vol 28

Feng X, Yang J, Laine AF, Angelini ED (2017) Discriminative localization in CNNS for weakly-supervised segmentation of pulmonary nodules. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 568–576

Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR (2012) Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580

Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on machine learning, pp 448–456

Jacobs C, van Rikxoort EM, Twellmann T, Scholten ET, de Jong PA, Kuhnigk JM, Oudkerk M, de Koning HJ, Prokop M, Schaefer-Prokop C et al (2014) Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. Med Image Anal 18(2):374–384

Kuhnigk JM, Dicken V, Bornemann L, Bakai A, Wormanns D, Krass S, Peitgen HO (2006) Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic ct scans. IEEE Trans Med Imaging 25(4):417–434

Lin TY, Goyal P, Girshick R, He K, Dollar P (2017) Focal loss for dense object detection. In: Proceedings of the ieee conference on computer vision and pattern recognition, pp 2980–2988

Messay T, Hardie RC, Rogers SK (2010) A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal 14(3):390–406

Mingliang X, Pei L, Mingyuan L, Hao F, Hongling Z, Bing Z, Yusong L, Liwei Z (2016) Medical image denoising by parallel non-local means. Neurocomputing 195:117–122

Murphy K, van Ginneken B, Schilham AM, De Hoop B, Gietema H, Prokop M (2009) A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. Med Image Anal 13(5):757–770

Niemeijer M, Loog M, Abramoff MD, Viergever MA, Prokop M, van Ginneken B (2011) On combining computer-aided detection systems. IEEE Trans Med Imaging 30(2):215–223

Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M (2013) Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, Heidelberg, pp 246–253

Ramaswamy S, Truong K (2017) Pulmonary nodule classification with convolutional neural networks. http://cs231n.stanford.edu/reports/2016/pdfs/324_Report.pdf

Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 234–241

Setio AA, Jacobs C, Gelderblom J, Ginneken B (2015) Automatic detection of large pulmonary solid nodules in thoracic CT images. Med Phys 42(10):5642–5653

Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, Wille MMW, Naqibullah M, Sánchez CI, van Ginneken B (2016) Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging 35(5):1160–1169

Shao L, Wu D, Li X (2014) Learning deep and wide: a spectral method for learning deep networks. IEEE Trans Neural Netw Learn Syst 25(12):2303–2308

Sirinukunwattana K, Raza SEA, Tsang YW, Snead DR, Cree IA, Rajpoot NM (2016) Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans Med Imaging 35(5):1196–1206

Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 35(5):1299–1312

Tang W, Wang Y, He W (2016) An image segmentation algorithm based on improved multiscale random field model in wavelet domain. J Ambient Intell Humaniz Comput 7(2):221–228

Van Ginneken B, Setio AA, Jacobs C, Ciompi F (2015) Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: Biomedical imaging (ISBI), 2015 IEEE 12th international symposium. IEEE, pp 286–289

Wu D, Pigou L, Kindermans PJ, Le NDH, Shao L, Dambre J, Odobez JM (2016) Deep dynamic neural networks for multimodal gesture segmentation and recognition. IEEE Trans Pattern Anal Mach Intell 38(8):1583–1597

Yang H, Yu H, Wang G (2016) Deep learning for the classification of lung nodules. arXiv:1611.06651