



OPEN ACCESS

EDITED BY

Omid Haji Maghsoudi,
Tempus Labs, United States

REVIEWED BY

Lichi Zhang,
Shanghai Jiao Tong University, China,
Mehul Sampat,
Flywheel, United States

*CORRESPONDENCE

Yingli Tian
ytian@ccny.cuny.edu

SPECIALTY SECTION

This article was submitted to Artificial
Intelligence in Radiology, a section of the
journal Frontiers in Radiology

RECEIVED 11 September 2022

ACCEPTED 28 November 2022

PUBLISHED 15 December 2022

CITATION

Liu J, Cao L, Akin O and Tian Y (2022) Robust
and accurate pulmonary nodule detection with
self-supervised feature learning on domain
adaptation.
Front. Radio 2:1041518.
doi: 10.3389/fradi.2022.1041518

COPYRIGHT

© 2022 Liu, Cao, Akin and Tian. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Robust and accurate pulmonary nodule detection with self-supervised feature learning on domain adaptation

Jingya Liu¹, Liangliang Cao², Oguz Akin³ and Yingli Tian^{1*}

¹The City College of New York, New York, NY, USA, ²UMass CI, Amherst, MA, USA, ³Memorial Sloan Kettering Cancer Center, New York, NY, USA

Medical imaging data annotation is expensive and time-consuming. Supervised deep learning approaches may encounter overfitting if trained with limited medical data, and further affect the robustness of computer-aided diagnosis (CAD) on CT scans collected by various scanner vendors. Additionally, the high false-positive rate in automatic lung nodule detection methods prevents their applications in daily clinical routine diagnosis. To tackle these issues, we first introduce a novel self-learning schema to train a pre-trained model by learning rich feature representatives from large-scale unlabeled data without extra annotation, which guarantees a consistent detection performance over novel datasets. Then, a 3D feature pyramid network (3DFPN) is proposed for high-sensitivity nodule detection by extracting multi-scale features, where the weights of the backbone network are initialized by the pre-trained model and then fine-tuned in a supervised manner. Further, a High Sensitivity and Specificity (HS^2) network is proposed to reduce false positives by tracking the appearance changes among continuous CT slices on Location History Images (LHI) for the detected nodule candidates. The proposed method's performance and robustness are evaluated on several publicly available datasets, including LUNA16, SPIE-AAPM, LungTIME, and HMS. Our proposed detector achieves the state-of-the-art result of 90.6% sensitivity at 1/8 false positive per scan on the LUNA16 dataset. The proposed framework's generalizability has been evaluated on three additional datasets (i.e., SPIE-AAPM, LungTIME, and HMS) captured by different types of CT scanners.

KEYWORDS

self-supervised learning, lung nodule detection, false positive reduction, feature pyramid network, medical image analysis, deep learning, feature learning

1. Introduction

Lung cancer is one of the world's leading cancers in terms of incidence and mortality rates (1). At the time of diagnosis, the disease stage is closely related to the survival of patients with lung cancer. Therefore, it is critical for the efforts to identify and intervene in lung cancer in the early stage (2). Computed tomography (CT) has been shown to visualize tumors better in early clinical diagnosis (3). However, it is cumbersome and time-consuming for radiologists to detect and label tumors on CTs

manually. To better assist the diagnosis of lung cancer, CT-based automatic pulmonary nodule detection methods have been widely explored (4–9) to develop computer-aided diagnosis (CAD) systems (10,11). The CAD framework for nodule detection commonly consists of a nodule detector identifying and detecting the location of nodule candidates and a classifier further distinguishing the false detected candidates from true nodules with a false positive reduction procedure. In recent years, deep learning based methods demonstrated excellent performance on medical image analysis and preliminary work of lung nodule detection (12–16) yielded a high sensitivity of over 95% on LUNA16 challenge dataset (17), but at a high false positive rate (i.e., 8 false positives per scan) which limited their uses in real clinical processes. Reducing the false-positive rate remains an open question. Most of the existing methods achieved sensitivities below 75% at 1/8 false positive per scan. The high false-positive rate is mainly caused by the following two reasons. (1) Some normal tissues are morphologically similar to the nodules in the CT image, leading to a high false detection rate. The approach for differentiating between the tissue and nodule is very crucial to reduce false positives for automatic lung nodule detection scheme. (2) The volumes of nodules significantly differ from the total CT volume, which may lead to the miss detection of some nodules. For example, in the LUNA16 dataset (17), nodules size can range from 3 mm to 30 mm (diameter), which varies up to 10 times. Only 0.059% of the total CT scan volume is occupied by a 10 mm nodule in diameter on a CT scan in a resolution of 213×293 pixels and 281 slices. Therefore, it is essential to design methods for detecting small nodules from large CT scan volume and further distinguish the normal tissues with similar appearances of nodules for CT scans with various machine settings and intensity scales.

For a data-dependent deep learning-based framework, artifacts such as intensity, machine setting, machine noise, and image protocol for collected CT scans could cause systematical differences. In order to develop robust deep learning-based lung nodule detection methods to handle CTs collected by different vendors of CT scanners, there is a need for a large amount of training data to be labeled. However, manually annotating a large number of CT scans can be tedious, attention-demanding, and time-consuming. It also requires human expertise in the specialty of radiology. Recently, a self-supervised learning approach (18,19) is proposed to learn the intermediate representation from the sizeable unlabeled dataset by a well-designed pretext task in a supervised learning manner. Inspired by the rotation ConvNets (20,21), in this paper, we simply rotate the CT scan at certain angles and design a rotation classification network as the pretext task to distinguish the rotation angle of each CT scan. The well-trained model effectively learns the rich features and semantic concept of the CT scans from a large unlabeled dataset and is then further

applied as a pre-trained model of the nodule detector for robust nodule detection training on small annotated datasets.

In this paper, as shown in Figure 1, our proposed framework contains three main components. (1) To improve the robustness of the nodule detector across datasets without requiring extra annotations, a self-pre-trained model is proposed to learn the rich spatial features among CT scans obtained from various manufacturers and obtained through the training by applying rotation prediction as pretext task. (2) Motivated by 2D Feature Pyramid Network (FPN) (22), a 3D high-sensitivity feature pyramid network (3DFPN) is developed for multi-scale feature prediction by combining the low-level high-resolution feature with high-level semantic features for the nodules with different sizes. (3) A false positive reduction network based on location history images to distinguish differences in the spatial distribution of nodules and normal tissues in continuous CT slices significantly eliminates the false positives while maintaining high sensitivity and specificity.

We proposed an accurate and robust pulmonary nodule detection framework (3DFPN-HS²) by integrating an accurate nodule detection model with a novel false positive reduction method to achieve the high sensitivity and specificity of diagnosis. This paper is an extension of our preliminary work (23), and the new contributions are summarized as follows:

1. To improve the robustness and generality of the nodule detector without additional annotations, we adopted a pre-trained model that can significantly improve the performance of the model across datasets with a simple and effective self-supervised learning schema.
2. By combining the pre-trained model with the two-stage framework (3DFPN-HS²), the experiments and results on the LUNA16 dataset demonstrate state-of-the-art performance, especially at low false-positive rates.
3. The generalizability of the proposed framework has been evaluated on three additional small datasets (SPIE-AAPM, LungTIME, and HMS) captured by different types of CT scanners show the robustness of the proposed framework, which has a high potential for application in clinical practice.

The remainder of this paper is organized as follows. Section 2 introduces the related work on self-supervised feature learning, object detection, and lung nodule detection from CT scans. Section 3 explains the proposed method. Section 4 presents the implementation details, experimental results, and discussions. Finally, Section 5 summarizes the remarks of this paper.

2. Related work

As data-driven computational mechanisms, supervised convolutional neural networks (ConvNets) usually require large-scale labeled data to obtain good performance and

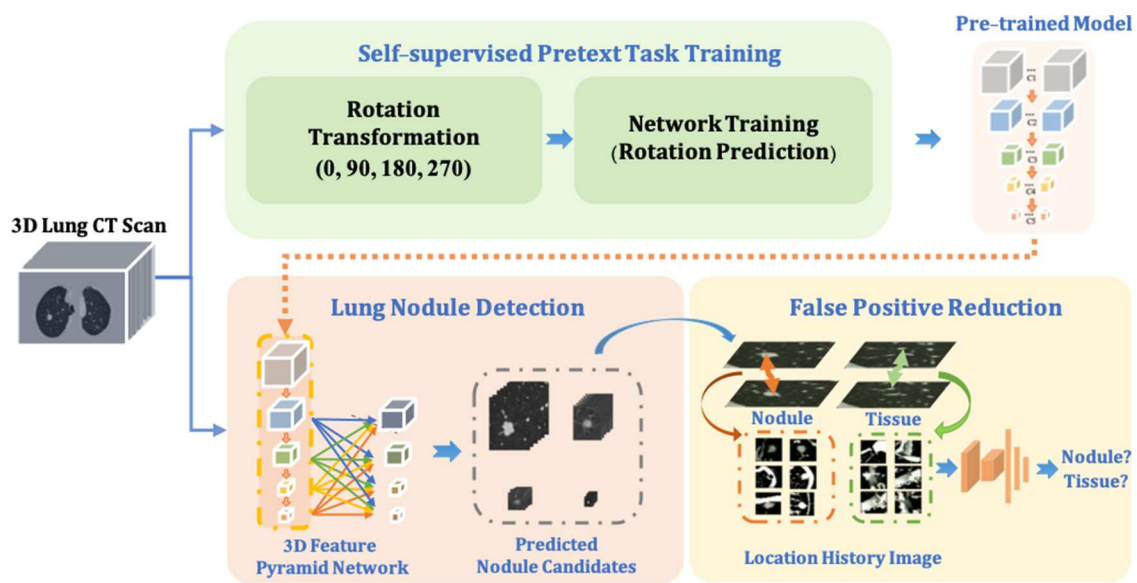


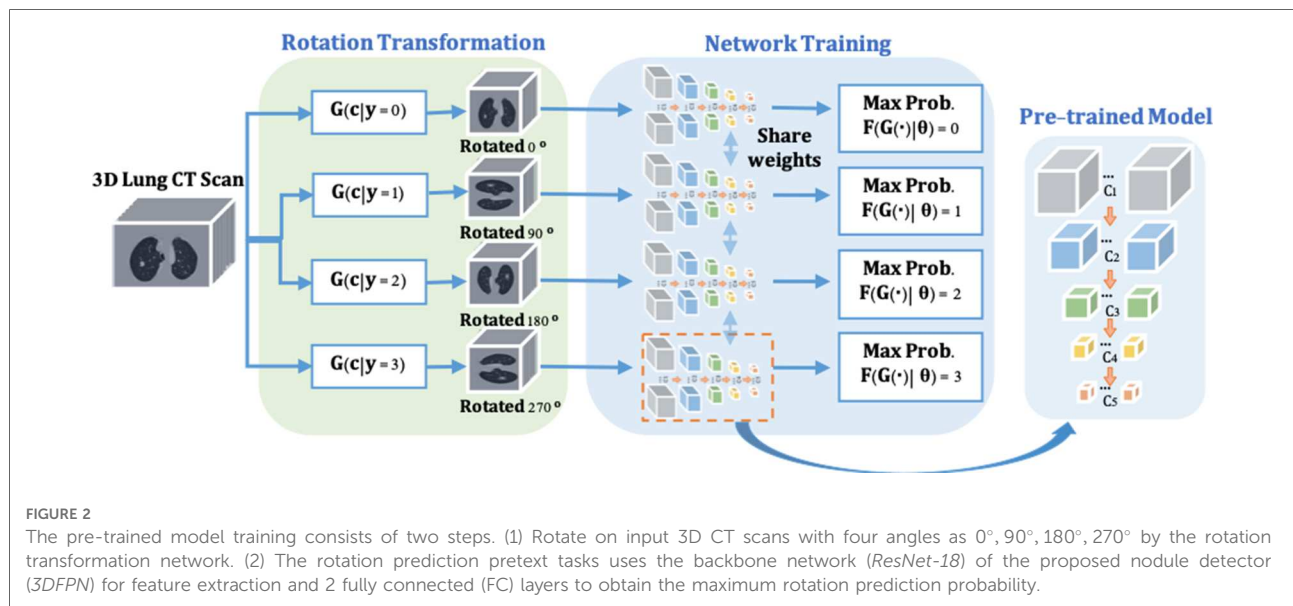
FIGURE 1

The proposed robust nodule detection framework $3DFPN-HS^2$ consists of 3D Feature Pyramid ConvNet ($3DFPN$) as lung nodule detector and HS^2 as false positive reduction network for high sensitivity and specificity lung nodule detection. (1) To improve the robustness of the framework across different datasets, a pre-trained model trained on the backbone network of detector *ResNet-18* is applied to the nodule detection network. The pre-trained model is obtained by a simple yet effective pretext task training through a rotation prediction network. The original CTs are rotated through a geometric transformation at (0, 90, 180, 270) degrees and followed by a classification network to predict the rotated angles of CT scans. (2) The $3DFPN$ takes the input of the entire CT scan to predict nodule candidates. The backbone network (*ResNet-18*) of $3DFPN$ is initialized by the weights from the pretrained model and then is fine-tuned with small datasets with annotation for pulmonary nodule detection in a supervised schema. (3) For the detected nodule candidates, the HS^2 network eliminates the false prediction of normal tissues based on the change in position of the continuous CT slices on LHI images. The detailed structure of self-supervised pretext task training is shown in Figure 2, the proposed $3DFPN$ network can be found in Figure 3, and LHI is illustrated in Figure 4.

overcome overfitting. Manually labeling a large number of CT scans is very expensive and requires multiple expert radiologists to perform the task to address reader agreement and variability issues. Therefore, in computer vision, some researchers proposed self- or un-supervised learning methods to learn feature representations without requiring manual data annotations (24–28). The intermediate representations of images and videos are learned by training the networks on one or multiple pretext tasks (e.g., regression or classification) with the modification of unlabeled data.

Recently, self-supervised learning methods are widely explored, and various pretext tasks have been proposed with learning by distinguishing the distorted transformations (18), adopting patches to predict relative position (29), coloring to map the image to a distribution (30), and distinguishing jigsaw puzzle with shuffle patches (19). Zhuang *et al.* (31) proposed a Rubik's cube task which extended jigsaw puzzle (19) of re-ordering 2D image patches to rotate and re-order 3D cubes. The result demonstrated the performance improvement for classification and segmentation tasks on CTs. Zhou *et al.* (32) proposed model genesis as the pretext task training through image distortion, in-painting, and unified method and have proven to benefit the downstream tasks on image classification and segmentation without any

annotation. Jing and Tian reviewed self-supervised learning methods in the comprehensive survey paper (33). The previous work of self-supervised learning related to lung nodule mainly focuses on nodule classification and segmentation. In this paper, we aim to demonstrate the robustness of the self-supervised learning method on the pulmonary nodule detection task with the rich semantic features learned from large scale lung CT scans. (20) rotated each input image by a multiple of 90 degrees and learned the semantic content of the image through an image rotation prediction network. However, previous methods are based on two dimensions and lacked spatial information. Recently, Jing and Tian (21) designed a rotation transformation network for a 3D input sequence to learn rich features from the video. The network can learn high-level spatial information of objects in videos while predicting the correct rotations. Following the framework in (21), by treating each CT scan as a video, we employ simple yet effective rotation prediction pretext task for predicting the rotation angle of 3D CT scans to obtain rich spatial information of CT scans. Several deep learning-based frameworks for object detection are proposed to handle small-scale and multi-scale objects (34,35). Single Shot multi-box Detector (SSD) (36) applied the pyramid feature hierarchy in the deep convolutional network, which



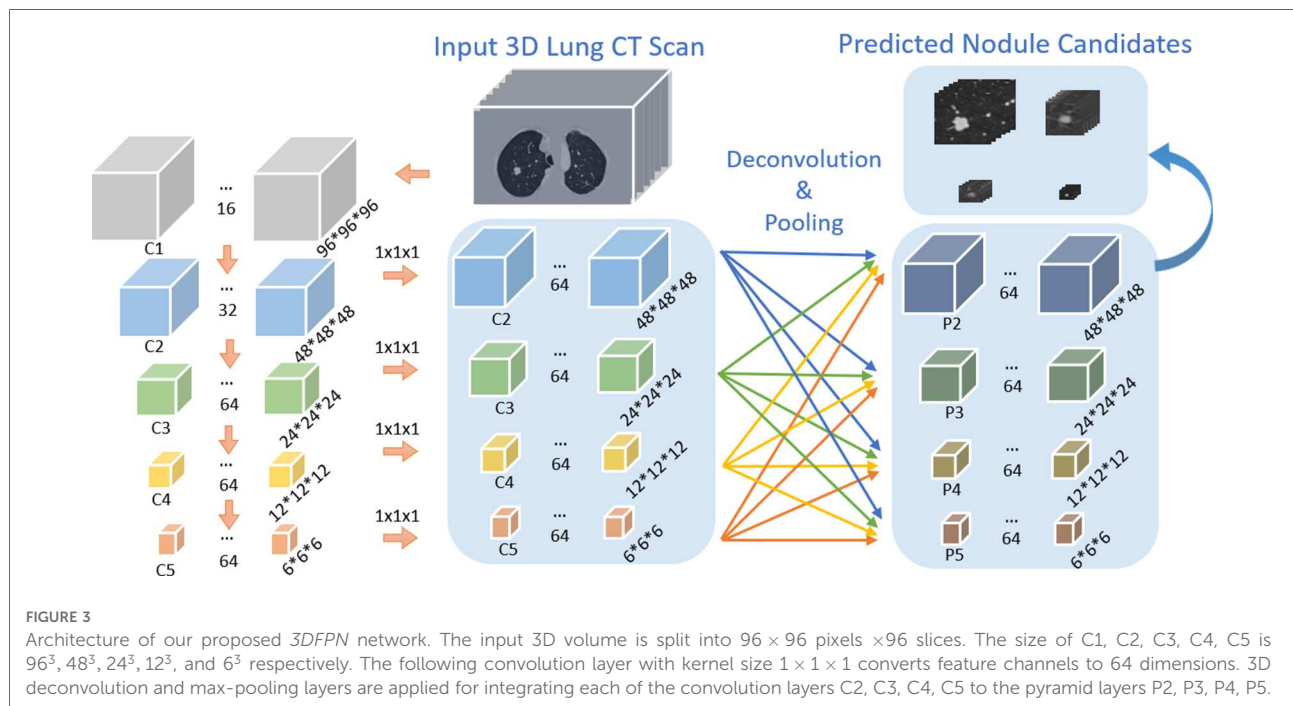
directly detected multi-scale objects by using multi-layer feature mapping in a single pass. However, *SSD* cannot reuse low-level feature maps that cause the miss detection of small objects. In order to detect small objects, *Scale Normalization for Image Pyramids (SNIP)* (35) selectively back-propagated the gradient of objects in different scales. Although small object detection performance has been significantly improved, the computation cost could be very high by applying multiple images as input. To date, a 2D Feature Pyramid Network (*FPN*) (22) demonstrated the effectiveness of small object detection by extracting the multi-scale feature maps containing the general low-level features of objects at different scales. A top-down path was introduced to pass global context information through lateral connections of high-level and low-level features. The computation of feature extraction was reduced by directly applying multi-scale feature maps. This *FPN* framework can be applied to lung nodule detection on each 2D CT slice. However, without 3D information among CT slices, high false positives were produced.

Continuous efforts have been made to detect pulmonary nodules with CT scans. Compared with the traditional methods based on intensity, shape, texture features, and context features, the deep learning-based methods have shown significant performance improvements (37–40). (16) could achieve an average sensitivity of 84.2% by a *3D Faster R-CNN* detector to learn the rich nodule features combined with a dual-path network and an encoder-decoder structure without false positive reduction, while *Single-Shot Single-Scale Lung Nodule Detection (S4ND)* (14) introduced 3D dense connections and investigated a down-sampling method for small nodule detection. These frameworks employed only a single scale feature map and were limited in the detection of nodules with a large range of sizes. (41)

proposed a multi-scale nodule detection method. The method firstly segmented the lung boundary delineation to obtain the lung region and then applied three sub-algorithms to detect the candidates in the three nodule size intervals. Although nodules of different sizes were treated separately, the sensitivity to 85.6% at 8 false positives per scan was limited by the rule-based threshold and morphological algorithms.

Furthermore, in order to reduce false positives, Dou et al. (13,42) applied three different 3D ConvNet architectures to adapt the different scales of nodules, and manually set the threshold to combine the weights. Ding et al. (12) applied a framework of *2D Faster R-CNN* nodule detector with a false-positive reduction classifier and obtained 89.1% average sensitivity. Multi-scale Gradual Integration Convolutional Neural Network (*MGI-CNN*) (43) used the image pyramid network with a multi-stream feature integration for small nodule detection and false-positive reduction. However, the computation cost of these frameworks is expensive due to the great effort to extract feature maps from images in different sizes and the multiple training process. Wang et al. (15) applied a *2DFPN* network for lung nodule detection, followed by a Conditional 3-Dimensional Non-Maximum Suppression (Conditional *3D-NMS*) and Attention 3D CNN (Attention *3D-CNN*) for false-positive reduction. However, without the spatial features within continuous CT slices, the high false-positive candidates were introduced, resulting in great efforts in the reduction process.

In this paper, we propose a rich spatial feature extraction method, an accurate multi-scale nodule detection network, and an efficient false-positive reduction algorithm for accurate and robust pulmonary nodule detection.



3. Method

In this section, we describe the details of the proposed accurate and robust pulmonary nodule detection framework. As shown in **Figure 1**, a pre-trained model is firstly obtained by employing the rotation prediction as the pretext task to extract rich spatial features and leverage the noises of CT scans captured by different manufacturers. The weights of the pre-trained model are applied to initialize the backbone network (*ResNet-18*) of the lung nodule detector 3DFPN. The 3DFPN takes an entire 3D volume of CT scan as input and outputs the 3D locations of lung nodule candidates. Then, the high sensitivity and specificity (HS^2) network predicts the probability of the true or false positive for the cropped 3D cube centered with the candidate nodules.

3.1. Self-supervised pre-trained model

To enable the proposed nodule detector for rich 3D CT feature extraction, a pre-trained model is obtained for the representative and discriminative features without using any additional labels. As shown in **Figure 2**, inspired by (20,21), a rotation transformation is first performed on 3D CT scans to obtain the rotation class with a certain angle. The rotation transform rotates 3D CT scans on the axial plane at an angle θ (0° , 90° , 180° , 270°). Predicting the correct rotation transformation of an image requires localizing orientations and types of salient objects. Classifying the rotation translation

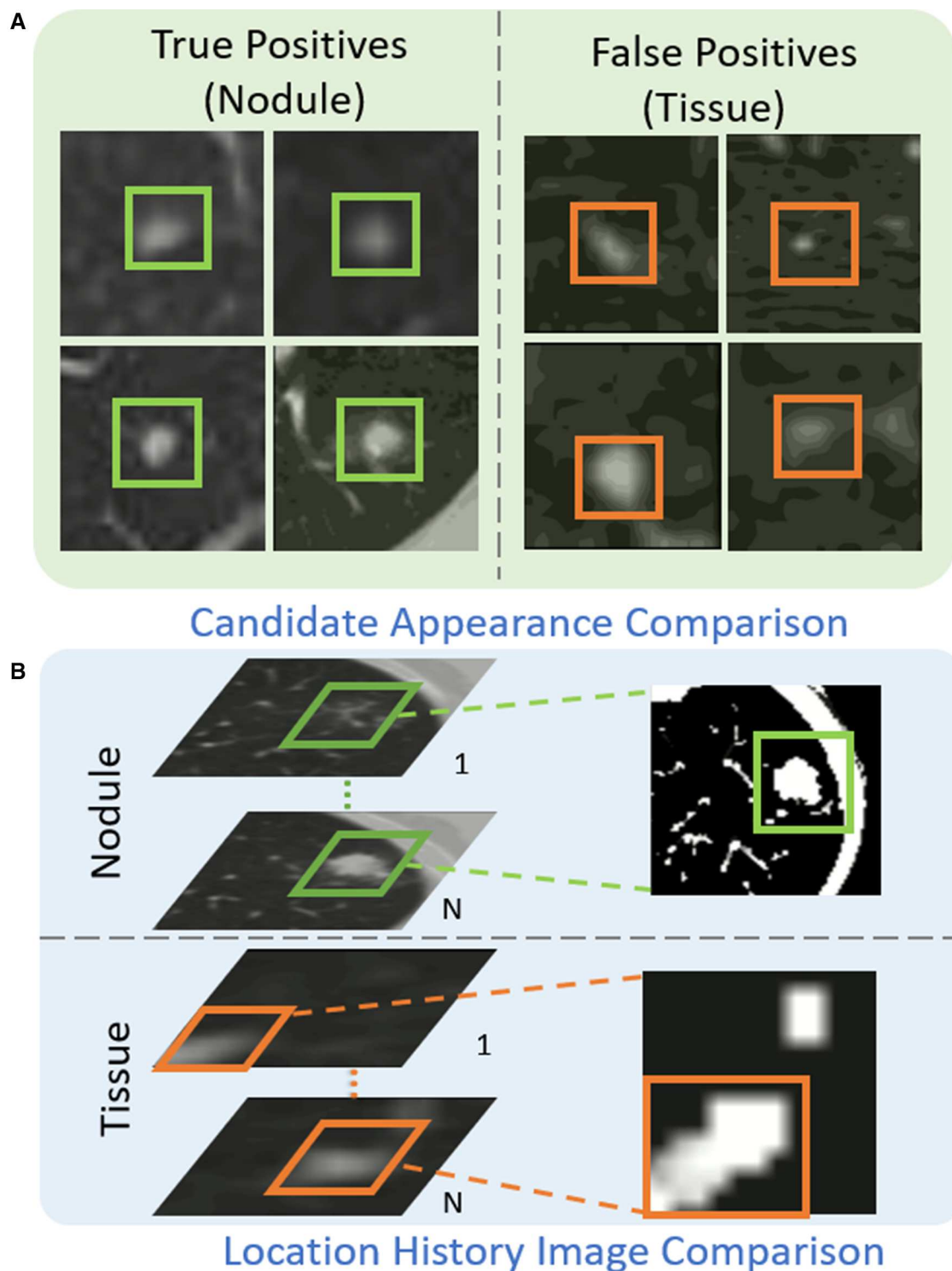
enables ConvNet to learn high-level spatial information of the objects. A dataset consisting of four rotating classes on CT scans is prepared for the pretext task (rotation prediction) training in a supervised manner, aiming to maximize the classification probability of rotation angle. The backbone network of nodule detector (*ResNet-18*) is applied to classify the rotation classes of the input CT scans, followed by two fully connected layers for probability prediction. Therefore, rich spatial CT scans features are learned by distinguishing the feature structures of the lung area in the CT scans. The cross-entropy loss is applied with K rotation angles (here $k = 4$) and rotation r as shown in Equation (1):

$$\text{loss}(c_i | \theta) = -1/K \sum_{r=1}^K \log(F(G(c_i, r) | \theta)), \quad (1)$$

where the classification network is defined as $F(\cdot | \theta)$ for spatial feature learning and the rotation transformation from the input 3D CT scans to the categories of rotation angle is represented as $G(c | y)$.

3.2. 3D feature pyramid network for nodule detection

The recent progress of computer vision suggests feature pyramid networks (FPN) for the powerful detection performance on objects at various scales (22). However, the original FPNs are designed to handle 2D images. Motivated

**FIGURE 4**

The proposed Location History Images (LHI) to distinguish tissues and nodules from the predicted nodule candidates. (A) The true nodules (green boxes) have similar appearances to the false detected tissues (orange boxes). (B) The location variances for nodules and tissues are oriented differently in LHIs. True nodules generally have a circular region representing the spatial changes as the brighter center (the size of the nodules decreases in the following CT slides) or a darker center (the size of the nodules increases in the following CT slides). On the other hand, the location variance for false detected tissues tends to change in certain directions, such as a gradual change of the trajectory line.

TABLE 1 Comparison between 2DFPN (22) and our proposed 3DFPN. With the 3D input of the proposed network, the feature pyramid layers are parallel connected with all the high and low-level features.

Method	Input 3D volume	Lateral connections	Integrate upper layer	Integrate lower layer	Upsample higher layer	Downsample lower layer
2DFPN (22)	No	✓	✓	No	✓	No
3DFPN (Ours)	✓	✓	✓	✓	✓	✓

by this, we propose a 3DFPN for 3D pulmonary nodule location detection from 3D CT volumetric scans. Different than (22), which only concatenating the upper-level features, to further preserve location details and obtain strong semantic features, a dense pyramid network is proposed by integrating the low-level and high-level layers for high-resolution and high-semantic features, respectively. Table 1 highlights the main differences between 2DFPN and our 3DFPN.

As shown in Figure 3, the bottom-up network extracts features from the convolution layers 2–5, refer to as C2, C3, C4, C5, followed by a convolution layer with kernel size $1 \times 1 \times 1$ to convert the features from the convolution layers to the same size. The feature pyramid network is composed of four layers: P2, P3, P4, P5. The max-pooling layer integrates low-level layer features with high-level features. 3DFPN predicts a confidence score and the corresponding location as $[x, y, z, d]$ for each nodule candidate, where $[x, y]$ are the spatial coordinates on each CT slice, z is the index number of the CT slice, d is the diameter of the nodule candidate. The backbone network of the detector is initialized with the weights of the pre-trained model and further refined with small labeled datasets with supervised learning.

3.3. HS² network for false positive reduction

As shown in Figure 4(A), the appearance of some tissues (orange box) is similar to that of real nodules (green box), which are likely to be detected as nodule candidates and generate a large number of false positives. Table 2 illustrates the analysis of 300 false positives predicted by the proposed nodule detector 3DFPN. We observe that 241 False Positives (FPs) are caused by the high appearance similarity of tissues (80.3%), 33 of them are due to inaccurate size detection (11%), and 26 FPs are caused by inaccurate location detection (8.7%). The majority of false positives are caused by normal tissue regions with a similar appearance. However, by treating each CT scan as a video, we discover that the orientations of

tissues and nodules present different patterns among the consecutive slices, as shown in Figs. 4(B), 5 and 6. The variance of true nodules tends to expand outward or diminish towards the center in continuous CT slices. Therefore, we propose a novel method to further distinguish the tissues among nodule candidates for false positive reduction.

Inspired by Motion History Image (MHI) (44,45), we define the Location History Image (LHI) as f . The intensity value of LHI within $(1, \tau)$ slice is represented by $f(x, y, s)$ by given any pixel location (x, y) on a CT slice s . The LHI is fed to a feed-forward neural network HS² with two convolutional layers and three fully connected layers. The HS² network refines predicted labels for true nodules and tissues.

The intensity of LHI is calculated according to Equation (2):

$$f(x, y, s) = \begin{cases} \tau & \text{if } \psi(x, y, s) = 1 \\ \max(0, f(x, y, s-1) - 1) & \text{otherwise} \end{cases}, \quad (2)$$

where the update function $\psi(x, y, s)$ is obtained by the spatial differentiation of the pixel intensity of two consecutive CT slices. The algorithm has the following steps. (1) If $|I(x, y, s) - I(x, y, s-1)|$ is larger than a threshold, $\psi(x, y, s) = 1$, otherwise, $\psi(x, y, s) = 0$. (2) For the current slice, if $\psi(x, y, s) = 1$, $f = \tau$. Otherwise, if $f(x, y, s)$ is not zero, it is attenuated with a gradient of 1. If $f(x, y, s)$ equals zero, then remains as zero. (3) Repeat steps (1) and (2) until all the slices are processed. Therefore, the proposed LHIs can adequately represent the location variance among continuous CT slices and their change patterns.

4. Experimental results and discussions

4.1. Datasets

In this paper, we employ training, testing, and performance evaluation on five public datasets: NLST, LUNA16, SPIE-AAPM, LungTime, and HMS Lung Cancer datasets. Table 3 summarizes the details of these datasets.

NLST Dataset: The national lung screening trial (NLST) (46) is a public dataset aimed to determine whether low-dose spiral CT screening for lung cancer can reduce lung cancer

TABLE 2 Statistic Analysis for False Positive Nodule Candidates.

	Tissue	Inaccurate Size	Inaccurate Location
Percentage	80.3%	11%	8.7%

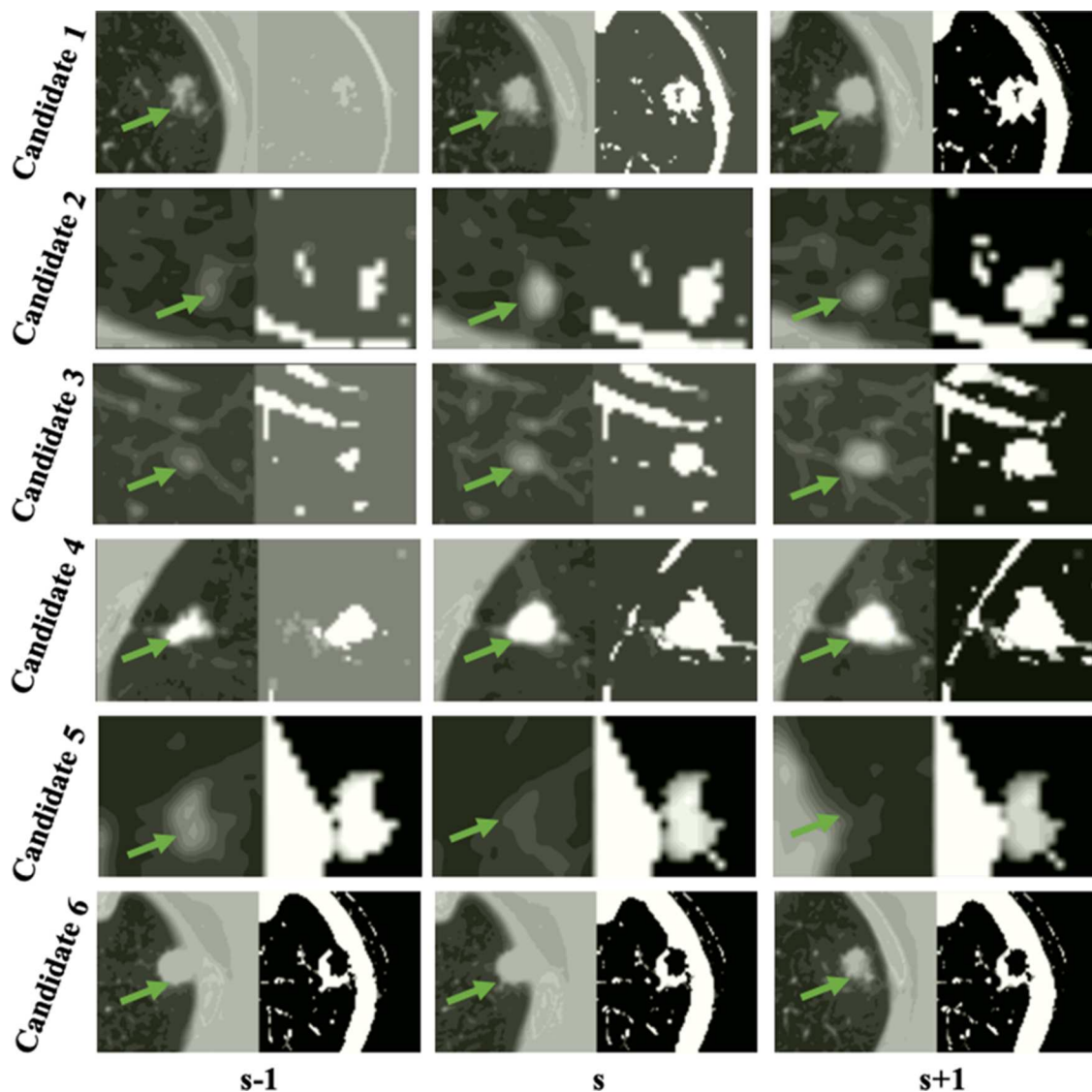


FIGURE 5

Examples of the detected true nodule candidates (the left image of each column) and their corresponding LHLs (the right image of each column) calculated between $(s-2, s-1)$, $(s-1, s)$, and $(s, s+1)$ slices shown in the $s-1$, s , $s+1$ columns. The green arrows mark the position of candidates. As shown in the figure, the true nodules have a circular region on LHL images as the location of the nodule approach to the center or the edge of nodule volume. Furthermore, the center location of the nodule candidates barely changes in the continuous slices.

mortality in high-risk populations compared with chest screening. The data includes participant characteristics, screening test results, diagnostic procedures, lung cancer, and mortality with more than 75,000 CT scans captured by four different manufacturers of CT scanners (i.e., GE, Philips, Siemens, and Toshiba.) Since this dataset is vast and there are no annotations for nodule locations, the NLST dataset is used for rotation transformation based self-supervised feature learning. In a total of 13,762 CT scans are applied for pretext task training.

LUNA16 Dataset: LUNA16 challenge dataset (17) contains 1,186 nodules, ranging in size from 3 to 30 mm

from 888 CT scans and agreed by at least 3 out of 4 radiologists. The dataset is officially divided into 10 subsets. To conduct a fair comparison with other lung nodule detection methods, we follow the same cross-validation protocol by applying 9 subsets as training and the remaining subset as testing and reporting the average performance. We split 10% of the training data used for validation to monitor the convergence of the training process. The nodule detector *3DFPN* is initialized by the pre-trained model trained on the NLST dataset and fine-tuned on the LUNA16 training subsets and perform the evaluation on the testing subset.

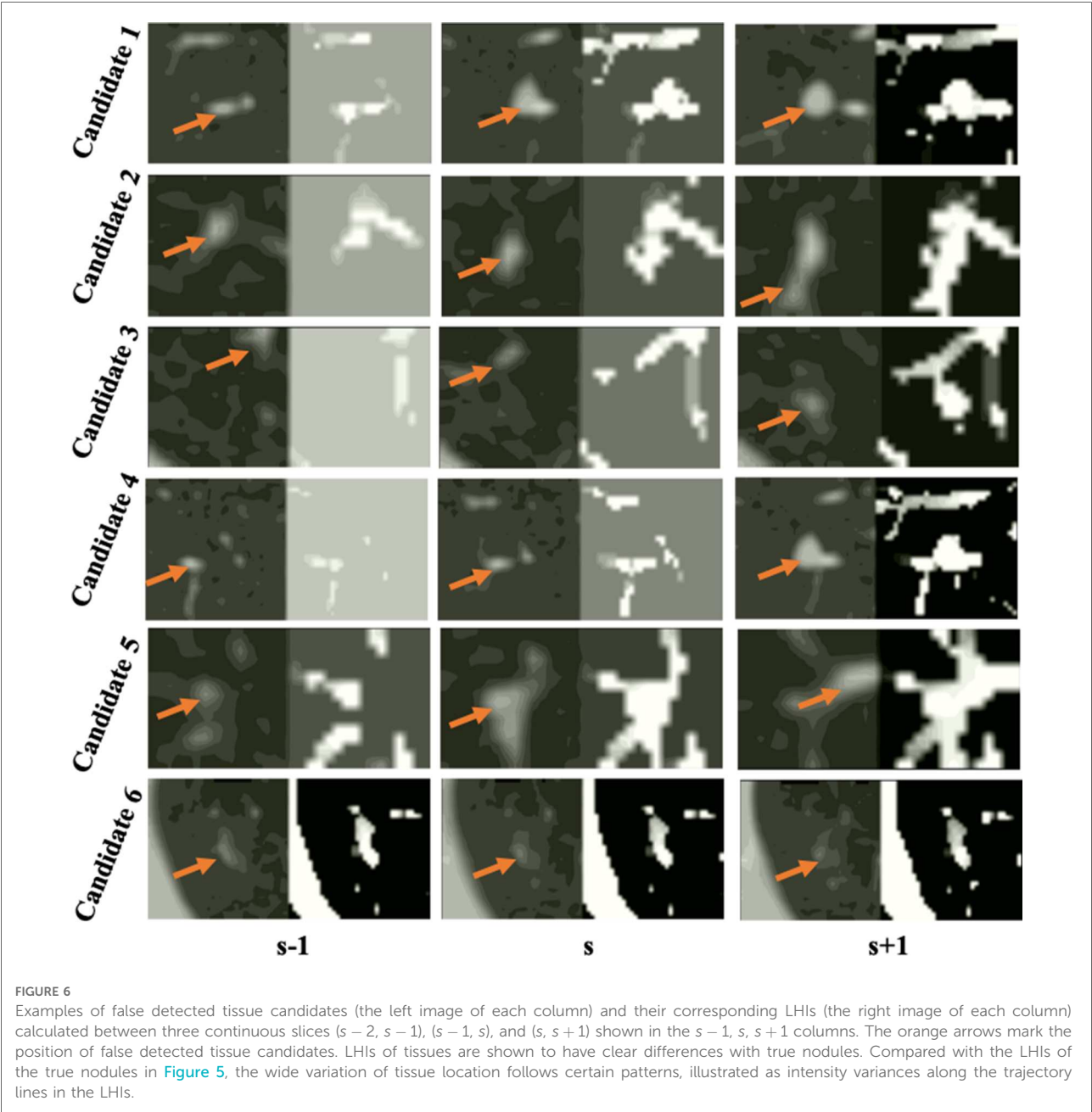


TABLE 3 Detailed Information of the Datasets.

Dataset	Manufacturer	CT number	Nodule number
NLST	GE Healthcare Philips Healthcare Siemens Healthcare Toshiba Healthcare	13,762	–
LUNA16	GE Healthcare	888	1186
SPIE-AAPM	Philips Healthcare	70	70
LungTIME	Siemens Healthcare	157	394
HMS Lung Cancer	GE Healthcare	229	254

SPIE-AAPM Dataset: The SPIE-AAPM dataset is collected for a 'Grand Challenge' of the diagnostic classification of malignant and benign lung nodule by the international society for optics and photonics (SPIE) with the support of American Association of Physicists in Medicine (AAPM) and the National Cancer Institute (NCI) (47). It contains 70 CT scans from 70 patients, with the annotation of nodule location and the nodule diagnosis categories of benign or malignant. It is applied in our paper for cross-dataset testing.

Lung TIME Dataset: The Lung Test Images from Motol Environment (Lung TIME) is publicly available and contains

157 CT scans with 394 nodules (48). The nodules are in the range of 2–10 mm in diameter. The CT scans annotations of the nodule location are provided. It is employed in our paper for cross-dataset testing.

HMS Lung Cancer Dataset: HMS Lung Cancer dataset (49) contains the CT scans and lung tumor sections generated by clinical care professionals used in competition with 461 patients. HMS contains a total number of 229 CT scans and 254 nodules with the nodule location annotation. It is employed in our paper for cross-dataset testing.

4.2. Data preprocessing

A preprocessing procedure is required to original CT scans for accurate nodule detection. First, the masks of the lung regions are extracted by lung region segmentation. The 2D single slice is processed first with a Gaussian filter to remove the fat, water, and kidney background and followed by a 3D connection volume extraction to remove unrelated areas (50). However, it takes 9 to 22 seconds to obtain the mask for each CT scan. To accelerate the processing speed for large datasets, we employ the LGAN method (51) and train the network on 10,000 CT slices for lung mask extraction to speed up the process in an average of 5 seconds per scan. Additionally, CT scans with a practical value of Hounsfield Unit between $[-1200, 600]$ are transformed into the gray value of $[0, 255]$ by a linear mapping. The spacing (mm/pixel) of CT scans between different patients and machines is various, and the re-sampling is applied to unify the spacing to 1 mm.

4.3. Experimental settings

4.3.1. Self-supervised pre-trained model

The 3D CT scans are rotated at four angles (0° , 90° , 180° , 270°). The backbone network of *3DFPN* (*ResNet-18*) is used to extract rich spatial features from the input CT scans, and two fully connected layers are applied to maximize the probability of rotation classes. The pre-trained model is then employed to initialize the weights for lung nodule detector *3DFPN*. During the training, the learning rate is set to 0.1, decreasing by $1/2$ after 70 and 85 epochs, with the weight decay of $5e^{-4}$. The total training includes 100 epochs, and the batch size is set to 16.

4.3.2. 3DFPN

The *3DFPN* network takes the entire CT scan as input and selects the volume of $96 \times 96 \times 96$ pixels through the sliding window schema. This size is selected experimentally to ensure that it accommodates the entire nodule even with the largest nodule (approx. 30 mm). In our *3DFPN*, the anchor sizes used to obtain candidate regions from the feature maps are $[3^3, 5^3, 10^3, 15^3, 20^3, 25^3, 30^3]$ pixels. The nodule positions

are predicted by 3D feature maps of the corresponding anchor regions. During the training process, the regions to the ground-truth regions with an Intersection-over-Union (IoU) threshold less than 0.02 are referred to as negative samples and with the threshold value greater than 0.4 are positive samples. To avoid the similarity between positive and negative samples, the regions between IoU values are ignored. We follow the *2DFPN* (22) to predict the nodule candidates with a 3×3 convolutional layer and followed by two 1×1 sibling convolution as classification and regression layer. The classification layer predicts the confidence score of the candidate classes, and the regression layer learns the offset between the region proposals and the ground-truth. Smooth $L1$ loss (52) and binary cross-entropy loss (BCE-loss) are used for location regression and classification, respectively. The proposals with a probability greater than 0.1 are selected as nodule candidates. Non-maximum suppression is further applied to eliminate multiple predicted nodule candidates.

4.3.3. HS^2 network

The HS^2 network consists of two convolution layers with 30 and 50 output channel dimensions, and three fully connected layers with (2048, 1024, 512) channel dimensions. The ReLU activation is applied after each convolution layer and followed by a batch normalization layer. 11 continuous CT slices are selected for LHI image generation with 5 slices before and after the current slice of nodule candidate. The sizes of the convolution kernel are set based on empirical experiments. Image patches are aligned with each predicted nodule candidate region but twice the size in both the x and y directions. To calculate the intensity of LHIs, the thresholds of spatial difference between two consecutive slices are set to 30 and 40 for data augmentation. LHIs are resized to 48×48 pixels as the input of the HS^2 network. To overcome the unbalanced data candidates, we randomly sample the false-positive candidates with a similar amount of true candidates and apply the data augmentation, including flipping, rotating, and cropping with 0.9 of the original size. In training, the learning rate starts at 0.01 and decreases to $1/10$ for every 500 epochs. 2,000 epochs are executed in training. The average prediction time for an entire CT scan is about 0.53 min/scan on one GeForce GTX 1080 GPU using Python 2.7.

4.4. Evaluation metrics

The performance is measured by Free-Response Receiver Operating Characteristic (FROC) analysis and Competition Performance Metric (CPM), the same as other methods. Following the LUNA16 challenge evaluation method, the FROC curve plots detection sensitivity and the points on the corresponding false positives curve are obtained by the true positive rate (true positives over the sum of true positives and

false negatives) while the false positive rate at 1/4, 1/2, 1, 2, 4, 8 per scan respectively. The CPM score is calculated by averaging the sensitivity of all false-positive levels per scan. Sensitivity is defined as the ratio of true positives divided by the total number of true positives and false negatives. Specificity is the ratio of true negatives over the total number of true negatives and false positives.

4.5. Experimental results and analysis

4.5.1. Comparison with other methods

Table 4 shows the FROC evaluation results at 1/8, 1/4, 1/2, 1, 2, 4, 8 false-positive levels for our proposed method compared with state-of-the-art methods on LUNA16. The highlighted numbers in the table represent the best performance for each column. Since most state-of-the-art methods do not use the pre-trained model, all the methods are tested on the LUNA16 dataset without the pre-trained model for a fair comparison, following the same FROC evaluation. The state-of-the-art methods are divided into two groups for the frameworks without false positive reduction (14,16,41) and with false positive reduction process (12,13,15,42,43). As shown in the table, compared with the state-of-the-art detection methods, the proposed *3DFPN* surpasses 13.9% false positive at 0.125 per scan and 2% on average CPM. Compared with the framework with the false positive reduction, our framework outperforms 5.5% average sensitivity over most results of other methods and 1% than Kim et al. (43). Moreover, the proposed framework achieves the best performance at most of FP levels. As mentioned above, CAD systems require not only high sensitivity but also high specificity. **Table 4** shows that the proposed *HS²* network significantly reduces false positives. *3DFPN-HS²* obtains the maximum sensitivity of 97.14% for

2 FPs per scan. Additionally, the proposed framework remains a high sensitivity above 90% for the 1/8, 1/4, and 1/2 FP per scan. The experimental results show that the proposed *3DFPN-HS²* reaches high sensitivity and specificity with state-of-the-art performance for lung nodule detection. The nodule detection results are shown in **Figure 7**.

To further analyze the performance of the detected network for various nodule sizes, we followed (53) to classify the test set into three categories. According to the size distribution of pulmonary nodules, the average CPM for 10-fold cross-validation are evaluated on the nodule sizes between 3 mm to 5 mm (small), 5 mm to 10 mm (medium), and larger than 10 mm (large), respectively. As shown in **Table 4**, *3DFPN-HS²* shows improvements in sensitivity for small, medium and large-sized nodule diameters compared with *3DFPN*. Specifically, the *3DFPN* performs well in detecting large and medium-sized nodule diameters as easy to detect. With the pre-trained model and false-positive reduction, the result shows a 3% improvement on average CPM for the small nodule detection and yields the best performance overall.

4.5.2. Robustness with the self-supervised pre-trained model

As shown in **Table 5**, we conduct two sets of experiments to assess the robustness of frameworks that does not apply the pretrained model and that employ the pretrained model. For the experiment without the pre-trained model, the *3DFPN-HS²* model is trained from scratch on the LUNA16 training set. The experiment with the pre-trained model applies the weights from the pre-trained model to initialize the model parameters, and then fine-tunes the model on the LUNA16 dataset, with the results highlighted in **Table 5**. The models are trained and fine-tuned only on LUNA 16 training set and further tested on

TABLE 4 FROC Performance comparison with the state-of-the-art methods on LUNA16 dataset: sensitivity and the corresponding false positives at 1/8, 1/4, 1/2, 1, 2, 4, 8 per scan. Our *3DFPN-HS²* method achieves the best performance (with >90% sensitivity) at most of false positive levels and significantly outperforms others especially at the low false positive levels (1/8 and 1/4). *3DFPN* indicates the nodule detector without false positive reduction. *3DFPN-HS²* is with false positive reduction. *3DFPN-HS² ★* shows the results applying pre-trained model trained on the NLST dataset. *CPM_s*, *CPM_m*, and *CPM_l* show the average detection performance for the small, medium, and larger nodules, respectively.

Methods	1/8	1/4	1/2	1	2	4	8	CPM Score	CPM _s	CPM _m	CPM _l
Without False Positive Reduction											
Gupta et al. (41)	0.531	0.629	0.790	0.835	0.843	0.848	0.856	0.763	–	–	–
Zhu et al. (16)	0.692	0.769	0.824	0.865	0.893	0.917	0.933	0.842	–	–	–
Khosravan et al. (14)	0.709	0.836	0.921	0.953	0.953	0.953	0.953	0.897	–	–	–
With False Positive Reduction											
Dou et al. (13)	0.677	0.737	0.815	0.848	0.879	0.907	0.922	0.827	–	–	–
Dou et al. (42)	0.659	0.745	0.819	0.865	0.906	0.933	0.946	0.839	–	–	–
Wang et al. (15)	0.676	0.776	0.879	0.949	0.958	0.958	0.958	0.878	–	–	–
Ding et al. (12)	0.748	0.853	0.887	0.922	0.938	0.944	0.946	0.891	–	–	–
Kim et al. (43)	0.904	0.931	0.943	0.947	0.952	0.956	0.962	0.942	–	–	–
3DFPN (Ours)	0.848	0.876	0.905	0.933	0.943	0.957	0.970	0.919	0.894	0.909	0.937
3DFPN-HS ² (Ours)	0.904	0.914	0.933	0.957	0.971	0.971	0.971	0.952	0.914	0.945	0.977
3DFPN-HS ² ★ (Ours)	0.906	0.923	0.948	0.981	0.981	0.981	0.981	0.957	0.928	0.953	0.979

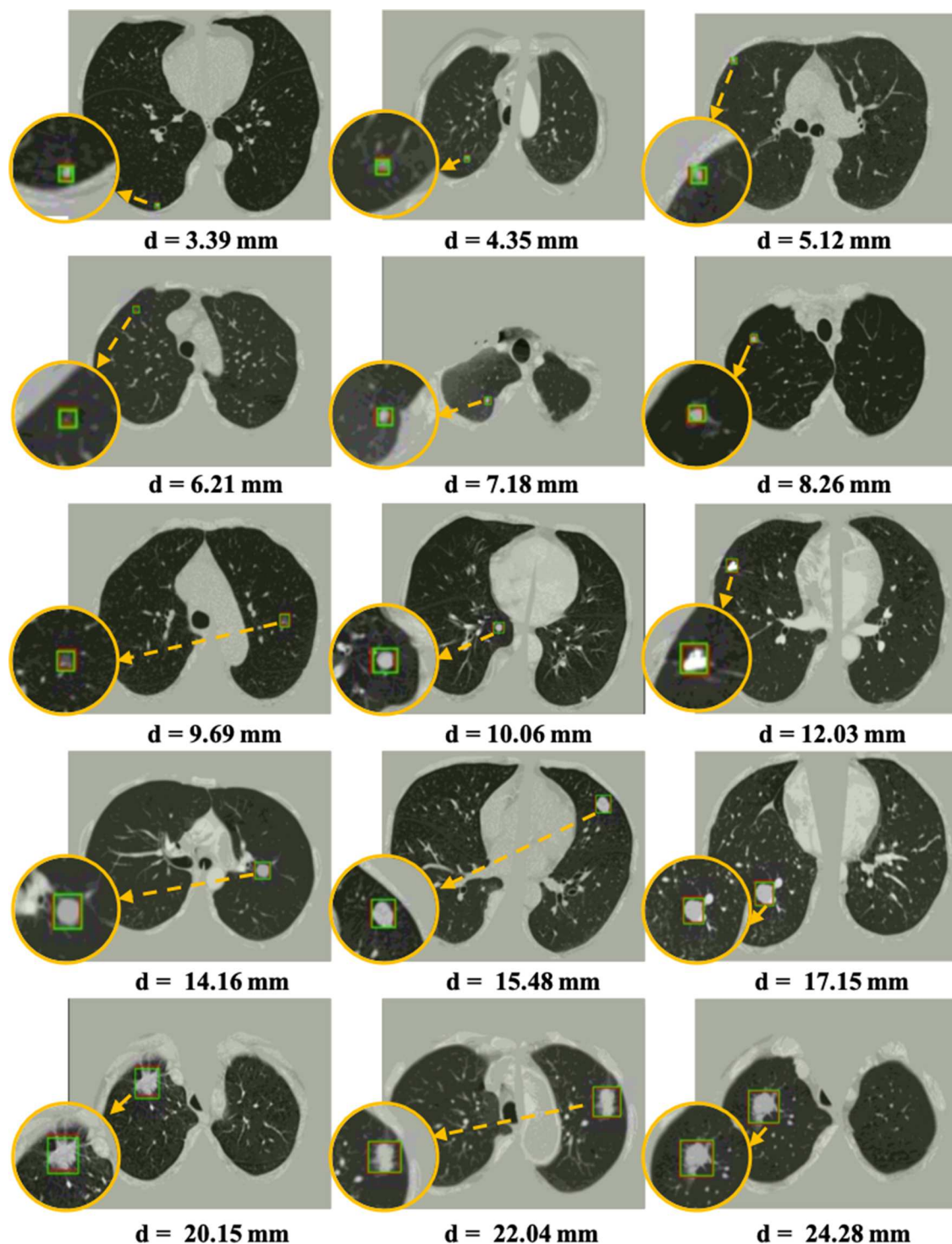


FIGURE 7

Visualization of some detected true nodules with different sizes from 3 mm to 25 mm in diameter d by our proposed $3DFPN-HS^2$ framework. For better visualization, the detected nodule regions are zoomed in, as shown in the orange circles. The green box indicates the predicted region, and the red box represents the ground-truth. Some of the red boxes are not observed because they are perfectly overlapped with the green boxes. The results demonstrate that our $3DFPN-HS^2$ framework is capable of detecting lung nodules of different sizes from CT scans accurately.

TABLE 5 FROC Performance comparison with and without using the pre-trained model, with and without false positive reduction: sensitivity and the corresponding false positives at 1/8, 1/4, 1/2, 1, 2, 4, 8 per scan. The *3DFPN-HS²* is only fine-tuned (with the pre-trained model)/trained (without the pre-trained model) on LUNA16 training set and tested on LUNA16 test set. At all levels of false positives per scan, the sensitivities of the framework using a pre-trained model and false positive reduction significantly outperform the one without the pre-trained model and false positive reduction.

Datasets	Pre-trained	HS^2	1/8	1/4	1/2	1	2	4	8	CPM Score
LUNA16	no	no	0.848	0.876	0.905	0.933	0.943	0.957	0.970	0.919
	no	✓	0.904	0.914	0.933	0.957	0.971	0.971	0.971	0.952
	✓	✓	0.906	0.923	0.948	0.981	0.981	0.981	0.981	0.957
SPIE-AAPM	no	no	0.764	0.791	0.839	0.873	0.895	0.914	0.923	0.857
	no	✓	0.808	0.852	0.877	0.914	0.928	0.947	0.964	0.899
	✓	✓	0.823	0.866	0.891	0.933	0.957	0.966	0.966	0.914
LungTIME	no	no	0.775	0.816	0.839	0.874	0.924	0.937	0.956	0.874
	no	✓	0.797	0.826	0.862	0.891	0.927	0.948	0.969	0.889
	✓	✓	0.814	0.837	0.874	0.899	0.931	0.952	0.974	0.897
HMS Lung Cancer	no	no	0.624	0.681	0.712	0.784	0.827	0.898	0.909	0.776
	no	✓	0.657	0.703	0.746	0.805	0.846	0.912	0.927	0.799
	✓	✓	0.752	0.823	0.856	0.894	0.921	0.941	0.958	0.877

the LUNA16 testing set and three different datasets (SPIE-AAPM, LungTime, and HMS Lung Cancer datasets) for cross-dataset validation. Additionally, we compare the experiments without and with the false positive reduction method. Using the pre-trained model shows slight improvements at all the false positive levels than those without the pre-trained model. Experiments performed on SPIE-AAPM, LungTime, and HMS Lung Cancer datasets with the nodule detector trained only on the LUNA16 dataset shows a significantly decreasing performance compared to the results on the LUNA16 testing set, especially for 1/8 false positive per scan. It is because LUNA16 has a relatively limited training set and cannot robust to other datasets. Compared with the model trained only on the LUNA16 dataset, the framework applying the self-supervised pre-trained model shows a significant improvement in all false positive levels on all of these datasets. Specifically, on 1/8 false positive per scan, the sensitivity is increased 7.4% on

SPIE-AAPM, 13.5% on LungTIME, and 8.9% on HMS, respectively. For the 8 false positives per scan, the accuracy is comparable to that of LUNA16. The significant improvement in performance demonstrates the robustness of applying the pre-trained model across different datasets without additional annotations. Because the model is trained on the LUNA16 training set, the test results on the LUNA16 test set have already achieved great performance as shown in [Table 5](#). Therefore, compared with the other three datasets, the performance of this model is only slightly improved on the LUNA16 test set with the pre-training model than without the pre-trained model.

4.5.3. Effectiveness of HS^2 network for FP reduction

The superiority of HS^2 network on LUNA16 dataset is demonstrated by two experiments. As shown in [Figure 8](#)

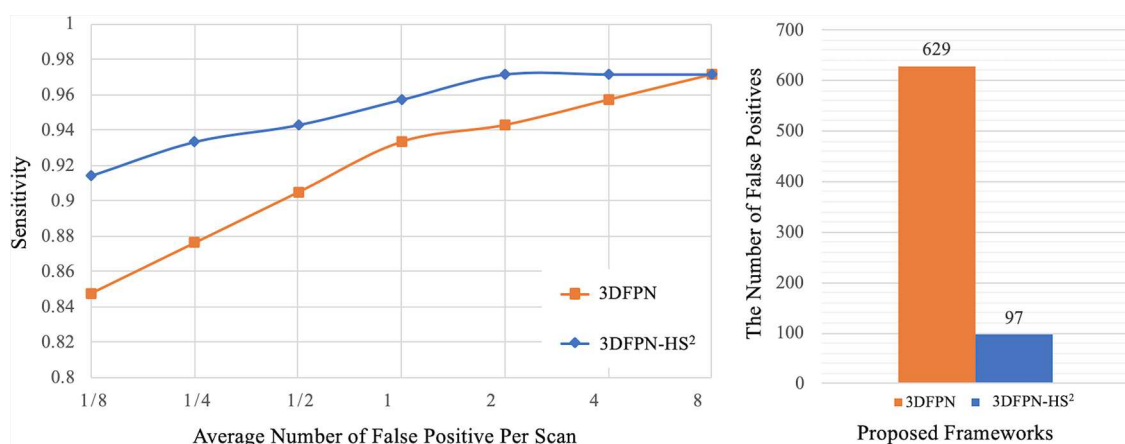


FIGURE 8

Comparison between the proposed *3DFPN* and *3DFPN-HS²*. Left: comparison of the proposed *3DFPN* and *3DFPN-HS²* on LUNA16 dataset without using the pre-trained model. *3DFPN-HS²* greatly improves the performance of the *3DFPN* at all FP levels per scan. Right: the number of false positives reduced from 629 to 97 for a total of 88 CT scans with the confidence score above 0 after applying the HS^2 network.

(left), the result of $3DFPN-HS^2$ with the false positive reduction is increased more than 5% at 1/8 FP level compared with $3DFPN$ without the HS^2 network. In addition, in Figure 8 (right), we further compare the numbers of FPs with (blue bar) and without (orange bar) HS^2 network in all the predicted nodule candidates in 88 CT scans (subset 9). The $3DFPN-HS^2$, with HS^2 for false positive reduction, can distinguish the false detected tissues from true nodules, significantly reducing FPs by 84.5%. In addition, our proposed $3DFPN$ without HS^2 network can still reach 97% at 8 FPs per scan, surpassing other state-of-the-art methods (shown in Table 4.)

5. Conclusion

In this paper, we have proposed an effective and robust $3DFPN-HS^2$ framework for lung nodule detection with a self-supervised feature learning schema. The different sizes of lung nodules can be detected by enriching the local and global features through a 3D feature pyramid network. By introducing the HS^2 network and treating each CT scan as a video, false positives are significantly reduced based on the patterns of location variance for nodules and tissues in continuous CT slices. Spatial features of CT scans can be effectively learned from large-scale CT scans without using additional labels by applying a self-supervised feature learning schema. The learned features can significantly improve the robustness of the proposed framework across different clinic datasets. As high sensitivity and specificity are achieved with robustness to the data from multiple CT scanner manufacturers, the proposed framework has a high potential in routine clinical practice.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* (2020) 70:7–30. doi: 10.3322/caac.21590
2. Hirsch FR, Franklin WA, Gazdar AF, Bunn PA. Early detection of lung cancer: clinical perspectives of recent advances in biology, radiology. *Clin Cancer Res.* (2001) 7:5–22.
3. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
4. Dou Q, Chen H, Yu L, Qin J, Heng P-A. Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection. *IEEE Trans Biomed Eng.* (2016) 64:1558–67. doi: 10.1109/TBME.2016.2613502
5. Huang X, Shan J, Vaidya V. Lung nodule detection in ct using 3d convolutional neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE (2017). p. 379–83.
6. Jiang H, Ma H, Qian W, Gao M, Li Y. An automatic detection system of lung nodule based on multigroup patch-based deep learning network. *IEEE J Biomed Health Inform.* (2017) 22:1227–37. doi: 10.1109/JBHI.2017.2725903
7. Li X, Shen L, Luo S. A solitary feature-based lung nodule detection approach for chest x-ray radiographs. *IEEE J Biomed Health Inform.* (2017) 22:516–24. doi: 10.1109/JBHI.2017.2661805
8. Ramachandran S, George J, Skaria S, Varun V. Using yolo based deep learning network for real time detection, localization of lung nodules from low

Ethics statement

Ethical review and approval was not required for this study in accordance with the local legislation and institutional requirements.

Author contributions

JYL, LLC, OA, and YLT contributed to the conception and design of the study. JYL wrote the first draft of the manuscript. JYL, LLC, OA, and YLT revised the sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version. All authors contributed to the article and approved the submitted version.

Funding

This material is based upon work supported by the National Science Foundation under award number IIS-2041307 and Memorial Sloan Kettering Cancer Center Support Grant/Core Grant P30 CA008748.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

dose ct scans. In *Medical Imaging 2018: Computer-Aided Diagnosis*. International Society for Optics and Photonics (2018). vol. 10575, p. 105751I.

9. Song W, Li S, Liu J, Qin H, Zhang B, Zhang S, et al. Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition. *IEEE J Biomed Health Inform.* (2018) 23:1215–24. doi: 10.1109/JBHI.2018.2852718

10. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005

11. Ohno Y, Koyama H, Yoshikawa T, Kishida Y, Seki S, Takenaka D, et al. Standard-, reduced-, and no-dose thin-section radiologic examinations: comparison of capability for nodule detection and nodule type assessment in patients suspected of having pulmonary nodules. *Radiology.* (2017) 284:562–73. doi: 10.1148/radiol.2017161037

12. Ding J, Li A, Hu Z, Wang L. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In *International Conference on Medical Image Computing, Computer-Assisted Intervention*. Springer (2017). p. 559–67.

13. Dou Q, Chen H, Yu L, Qin J, Heng P-A. Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection. *IEEE Trans Biomed Eng.* (2017) 64:1558–67. doi: 10.1109/TBME.2016.2613502

14. Khosravan N, Bagci U. S4nd: single-shot single-scale lung nodule detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2018). p. 794–802.

15. Wang B, Qi G, Tang S, Zhang L, Deng L, Zhang Y. Automated pulmonary nodule detection: high sensitivity with few candidates. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* Springer (2018). p. 759–67.

16. Zhu W, Liu C, Fan W, Xie X. Deeplung: deep 3D dual path nets for automated pulmonary nodule detection and classification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (2018). p. 673–81.

17. Aaa S, Traverso A, De BT, Msn B, Cvd B, Cerello P, et al. Validation, comparison, combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Med Image Anal.* (2017) 42:1–13. doi: 10.1016/j.media.2017.06.015

18. Dosovitskiy A, Fischer P, Springenberg JT, Riedmiller M, Brox T. Discriminative unsupervised feature learning with exemplar convolutional neural networks. *IEEE Trans Pattern Anal Mach Intell.* (2015) 38:1734–47. doi: 10.1109/TPAMI.2015.2496141

19. Noroozi M, Favaro P. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European Conference on Computer Vision*. Springer (2016). p. 69–84.

20. Gidaris S, Singh P, Komodakis N. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728* (2018).

21. Jing L, Tian Y. Self-supervised spatiotemporal feature learning by video geometric transformations. *arXiv preprint arXiv:1811.11387* (2018).

22. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017). p. 2117–25.

23. Liu J, Cao L, Akin O, Tian Y. 3DFPN-HS²: 3D feature pyramid network based high sensitivity and specificity pulmonary nodule detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2019).

24. Jing L, Tian Y. Self-supervised visual feature learning with deep neural networks: a survey. *arXiv preprint arXiv:1902.06162* (2019).

25. Mundhenk TN, Ho D, Chen BY. Improvements to context based self-supervised learning. In *Computer Vision and Pattern Recognition (CVPR)* (2018).

26. Nasrullah N, Sang J, Alam MS, Xiang H. Automated detection and classification for early stage lung cancer on ct images using deep learning. In *Pattern Recognition and Tracking XXX*. International Society for Optics and Photonics (2019). vol. 10995, p. 109950S.

27. Shariaty F, Mousavi M. Application of cad systems for the automatic detection of lung nodules. *Inform Med Unlocked.* (2019) 15:100173. doi: 10.1016/j.imu.2019.100173

28. Winkels M, Cohen TS. Pulmonary nodule detection in ct scans with equivariant cnns. *Med Image Anal.* (2019) 55:15–26. doi: 10.1016/j.media.2019.03.010

29. Doersch C, Gupta A, Efros AA. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE International Conference on Computer Vision* (2015). p. 1422–30.

30. Zhang R, Isola P, Efros AA. Colorful image colorization. In *European conference on computer vision* Springer (2016). p. 649–66.

31. Zhuang X, Li Y, Hu Y, Ma K, Yang Y, Zheng Y. Self-supervised feature learning for 3D medical images by playing a rubik's cube. In *International*

Conference on Medical Image Computing, Computer-Assisted Intervention. Springer (2019). p. 420–8.

32. Zhou Z, Sodha V, Siddiquee MMR, Feng R, Tajbakhsh N, Gotway MB, et al. Models genesis: generic autodidactic models for 3D medical image analysis. In *International Conference on Medical Image Computing, Computer-Assisted Intervention*. Springer (2019). p. 384–93.

33. Jing L, Tian Y. Self-supervised visual feature learning with deep neural networks: a survey. *IEEE Trans Pattern Anal Mach Intell.* (2020) 43:4037–58.

34. Jenuwine NM, Mahesh SN, Furst JD, Raicu DS. Lung nodule detection from ct scans using 3d convolutional neural networks without candidate selection. In *Medical Imaging 2018: Computer-Aided Diagnosis*. International Society for Optics and Photonics (2018). vol. 10575, p. 1057539.

35. Singh B, Davis LS. An analysis of scale invariance in object detection snip. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018). p. 3578–87.

36. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. Ssd: single shot multibox detector. In *European Conference on Computer Vision* (2016). p. 21–37.

37. Cao H, Liu H, Song E, Ma G, Jin R, Xu X, et al. A two-stage convolutional neural networks for lung nodule detection. *IEEE J Biomed Health Inform.* (2020) 24:2006–15.

38. Samala RK, Chan H -P, Richter C, Hadjiiski L, Zhou C, Wei J. Analysis of deep convolutional features for detection of lung nodules in computed tomography. In *Medical Imaging 2019: Computer-Aided Diagnosis*. International Society for Optics and Photonics (2019). vol. 10950, p. 109500Q.

39. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, Van Riel SJ, et al. Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging.* (2016) 35:1160–9. doi: 10.1109/TMI.2016.2536809

40. Zhu H, Zhao H, Song C, Bian Z, Bi Y, Liu T, et al. Mr-forest: a deep decision framework for false positive reduction in pulmonary nodule detection. *IEEE J Biomed Health Inform.* (2019) 24:1652–63.

41. Gupta A, Saar T, Martens O, Moullec YL. Automatic detection of multisize pulmonary nodules in ct images: large-scale validation of the false-positive reduction step. *Med Phys.* (2018) 45:1135–49. doi: 10.1002/mp.12746

42. Dou Q, Chen H, Jin Y, Lin H, Qin J, Heng P-A. Automated pulmonary nodule detection via 3d convnets with online sample filtering and hybrid-loss residual learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2017). p. 630–8.

43. Kim B-C, Yoon JS, Choi J-S, Suk H-I. Multi-scale gradual integration cnn for false positive reduction in pulmonary nodule detection. *Neural Netw.* (2019) 115:1–10.

44. Davis JW. Hierarchical motion history images for recognizing human motion. In *Proceedings IEEE Workshop on Detection, Recognition of Events in Video* (2001). p. 39–46.

45. Yang X, Zhang C, Tian Y. Recognizing actions using depth motion maps-based histograms of oriented gradients. In *Proceedings of the 20th ACM international conference on Multimedia* (2012). p. 1057–60.

46. National Lung Screening Trial Research Team. The national lung screening trial: overview and study design. *Radiology.* (2011) 258:243–53. doi: 10.1148/radiol.10091808

47. Armato III SG, Hadjiiski L, Tourassi G, Drukker K, Giger M, Li F, et al. Spie-aapm-nci lung nodule classification challenge dataset. *Cancer Imaging Arch.* (2015) 10:K9. doi: 10.1117/1.JMI.3.4.044506

48. Dolejsi M, Kybic J, Polovincak M, Tuma S. The lung time: annotated lung nodule dataset and nodule detection framework. In *Medical Imaging 2009: Computer-Aided Diagnosis*. International Society for Optics and Photonics (2009). vol. 7260, p. 72601U.

49. Mak RH, Endres MG, Paik JH, Sergeev RA, Aerts H, Williams CL, et al. Use of crowd innovation to develop an artificial intelligence-based solution for radiation therapy targeting. *JAMA Oncol.* (2019) 5:654–61. doi: 10.1001/jamaoncol.2019.0159

50. Liao F, Liang M, Li Z, Hu X, Song S. Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network. *IEEE Trans Neural Netw Learn Syst.* (2019) 30:3484–95. doi: 10.1109/TNNLS.2019.2892409

51. Tan J, Jing L, Huo Y, Tian Y, Akin O. Lgan: lung segmentation in ct scans using generative adversarial network. *arXiv preprint arXiv:1901.03473* (2019).

52. Girshick R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (2015). p. 1440–48.

53. Huang W, Hu L. Using a noisy u-net for detecting lung nodule candidates. *IEEE Access.* (2019) 7:67905–15. doi: 10.1109/ACCESS.2019.2918224