

## Article

# 3DAGNet: 3D Deep Attention and Global Search Network for Pulmonary Nodule Detection

Muwei Jian <sup>1,2,\*</sup>, Linsong Zhang <sup>1,†</sup>, Haodong Jin <sup>1</sup> and Xiaoguang Li <sup>3</sup>

<sup>1</sup> School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China; 212115031@mail.sdufe.edu.cn (L.Z.); jinhaodong@mail.sdufe.edu.cn (H.J.)

<sup>2</sup> School of Information Science and Technology, Linyi University, Linyi 276000, China

<sup>3</sup> Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; lxg@bjut.edu.cn

\* Correspondence: jianmuweihk@163.com

† These authors contributed equally to this work.

**Abstract:** In traditional clinical medicine, respiratory physicians or radiologists often identify the location of lung nodules by highlighting targets in consecutive CT slices, which is labor-intensive and easy-to-misdiagnose work. To achieve intelligent detection and diagnosis of CT lung nodules, we designed a 3D convolutional neural network, called 3DAGNet, for pulmonary nodule detection. Inspired by the diagnostic process of lung nodule localization by physicians, the 3DAGNet includes a spatial attention and a global search module. A multi-scale cascade module has also been introduced to enhance the model detection using attention enhancement, global information search, and contextual feature fusion. The experimental results showed that the proposed network achieved accurate detection of lung nodule information, and our method achieves a high sensitivity of 88.08% of the average FROC score on the LUNA16 dataset. In addition, ablation experiments also demonstrated the effectiveness of our method.

**Keywords:** intelligent detection; lung nodules; spatial attention and global search module; multi-scale cascade module



**Citation:** Jian, M.; Zhang, L.; Jin, H.; Li, X. 3DAGNet: 3D Deep Attention and Global Search Network for Pulmonary Nodule Detection. *Electronics* **2023**, *12*, 2333. <https://doi.org/10.3390/electronics12102333>

Academic Editors: Shaohui Lin, Fuhai Chen and Yunhang Shen

Received: 14 April 2023

Revised: 16 May 2023

Accepted: 17 May 2023

Published: 22 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Lung cancer is a malignant tumor formed by lesions in the bronchial mucosa or glands of the lung, which is one of the most life-threatening malignant tumors to human life. It has been at the head of the incidence and mortality rate among all cancers worldwide. The key to solving lung cancer lies in early and timely diagnosis and treatment. Lung cancer appears in the form of lung nodule lesions in the early stage. Therefore, the key to identifying malignant lesions and being able to treat lung cancer depends greatly on pinpointing the location of lung nodule lesions and identifying the characteristics of the lesions.

To diagnose pulmonary nodules early, chest computed tomography (CT), a widely used radiological method that produces detailed images of the lungs, has been well documented for its effectiveness. Nevertheless, CT images of a single patient often contain hundreds of slices, and it takes a lot of time and effort for even an experienced imaging physician to determine the location and characteristics of nodules from hundreds of slices. To relieve the stress and workload of physicians and to avoid subjective errors in diagnosis due to long working hours, computer-aided detection (CAD) systems have been designed to help physicians in lesion detection and analysis [1]. Recently developed CAD systems mainly focus on machine learning, especially deep learning, and use convolutional neural networks (CNNs) to simulate physicians' lesion diagnosis process to help physicians achieve lung nodule detection [2]. Zhao et al. [3] explored a novel agile CNN framework, based on the layer settings of LeNet and the parameter settings of AlexNet, to build a hybrid CNN model for lung nodule classification. Sori et al. [4] proposed a multi-path CNN for feature extraction of lung nodules using both local and global contextual features, and

finally merged the results of each branch to detect lung nodules. Tang et al. [5] designed an end-to-end 3D convolutional neural network to efficiently learn nodule information via online hard neg mining. The experimental results demonstrated that the model can effectively capture the spatial features of CT data for the detection and analysis of lung nodules. The above findings demonstrate the effectiveness of CNN in aiding diagnosis and can help physicians relieve the pressure of lung nodule detection. Although the existing lung nodule detection methods have achieved considerable progress, they still suffer from low sensitivity and a high false alarm rate.

Respiratory physicians or radiologists review entire pulmonary CT slices to identify lung nodules. The lung nodules can appear randomly in several consecutive slices, not throughout the entire CT image. To effectively address the specific challenge in nodule detection and simulate the way physicians locate and diagnose pulmonary nodules, we designed a 3D convolutional neural network with spatial and global attention enhancement modules for nodule detection.

The main contributions of this paper are summarized as follows:

- (1) A 3D convolutional neural network combining multi-scale feature fusion and an attention mechanism, named 3DAGNet, is proposed for the automatic detection of lung nodules, which achieve multi-scale feature fusion.
- (2) A two-branch attention module is designed to simulate the behavior of physicians' diagnosis. It considers the enhancement of attention to CT images from the depth direction in terms of the emergence of lung nodules, and the method combines the joint analysis of null convolution and attention mechanisms.
- (3) A three-branch multi-scale feature fusion module is designed to replace the process of decoding up-sampling in the traditional sense. The feature information at different scales is fused to explore both the high-level and low-level semantic features.
- (4) The results show that the proposed network significantly outperforms the compared state-of-the-art methods.

The remainder of this paper is organized as follows. In Section 2, the current status of related work that is closely related to this paper is presented. The combination of multi-scale feature fusion and attention mechanisms in our 3DAGNet is described in Section 3. In Section 4, the experimental results are presented and analyzed. Finally, the conclusions of the paper are provided in Section 5.

## 2. Related Work

### 2.1. Lung Nodule Detection

The detection of lung nodules using low-dose CT techniques is critical for the early detection of lung cancer [6]. However, this process can be quite burdensome for radiologists due to the large number of CT images that need to be reviewed. To address this problem, the advent of computer-aided detection (CAD) systems can greatly improve the efficiency of early screening for lung nodules, thus providing patients with more accurate diagnostic results. In recent years, the automatic detection of pulmonary nodules has received increasing attention, but it is still not easy to effectively reduce their false-positive rate. The reason is that, on the one hand, pulmonary nodules vary in shape, size, and type; on the other hand, some interstitial lung masses (e.g., blood vessels and pulmonary fibrosis) have a very similar appearance to real pulmonary nodules, which makes their accurate identification extremely difficult.

CAD, traditionally used to detect pulmonary nodules on chest CT, usually consists of two main stages: (i) the selection of candidate nodules (i.e., identification of nodules); and (ii) the removal of false-positive nodules (FPN) while retaining true-positive nodules (TPN), i.e., the classification of candidate nodules as nodules or non-nodules. False-positive nodes are usually excluded in the second stage. The most common classification methods employ feature-based classifiers [7]. First, the candidate nodules identified in the first step are segmented and then features are extracted in the segmented region. The extracted features usually include morphology-based features, gray-level-based features, and texture

features. There are many references of extracting specific desired features from segmented or non-segmented images first, and then feeding them into classifiers, such as support vector machines, decision trees, artificial neural networks, or integrated classifiers, to detect lung nodules. Several automatic lung nodule detection methods use traditional machine learning methods. Murphy et al. [8] designed a CAD system based on a K-Nearest Neighbor (KNN) classifier with a significantly lower false-positive rate. Lee et al. [9] used a random forest (RFF)-based CAD with a diagnostic sensitivity of 98.33% and a specificity of 97.11%, demonstrating the advantages of integrated learning in lung nodule detection. In order to deal with the imbalance issue between the number of nodules and non-nodules, Sui et al. [10] proposed a novel support vector machine classifier, i.e., RU-SMOTE-SVM classifier, to improve the performance. Traditional CAD systems based on texture and morphological evaluation have shown satisfactory results in the detection of lung nodules, but they tend to detect and analyze nodule features based on local features only from a statistical point of view, which is increasingly unable to meet the current requirements for high sensitivity and low false-positive rate in lung nodule detection.

Recently, deep convolutional neural networks have achieved great success in image processing [11–13] and have also been introduced into the field of medical imaging [14]. There have been many studies applying deep convolutional neural networks to computer-aided detection systems for pulmonary nodules. Li et al. [15] applied a method for pulmonary nodule detection using deep convolutional neural networks. Zhu et al. [16] introduced a 3D U-shaped residual network with the foundation of end-to-end detection and channel-wise attention mechanisms. The first step is the introduction of an upgraded attention gate (AG) that uses crucial feature dimensions at skip connections for feature propagation to lower the false-positive rate. To further increase detection sensitivity, a channel interaction unit (CIU) is created before the detection head. Additionally, the loss function for the gradient harmonization mechanism (GHM) is employed to address the issue of an imbalance between the positive and negative samples. Jin et al. [17] constructed a deep 3D residual convolutional neural network to reduce the false-positives of candidate nodules and the method achieved a high detection performance. Recently, Mei et al. [18] optimized the original non-local model in the channel dimension and designed a Slice Grouped Non-Local (SGNL) model that learns explicit correlations between any elements across slices. The SGNL module was combined with a 3D Region Proposal Network (RPN) with the aim of obtaining a high performance in the detection of nodules. The proposal network aims to obtain the long-range dependency between different dimensions. In addition, a large lung nodule classification dataset PN9 is proposed in the paper, with nine categories of lung nodules, which includes more than 8000 CT scans and more than 40,000 lung nodules. In the literature [19], Song et al. introduced a centroid matching network based on a 3D sphere representation. Their method consists of two components, namely sphere representation and centroid matching. First, to match nodule annotations in clinical practice, they propose a bounding sphere instead of the commonly used bounding box to represent nodules with center of mass, radius, and local offsets in three-dimensional space. A compatible spheroid-based intersecting excess joint loss function is introduced to train the lung nodule detection network stably and efficiently. Secondly, the method naturally discards the pre-determined anchor box by designing an aggressive center-points selection and matching process to make the network anchor-free. A significant advantage of CNN is that it does not require any feature extraction from the image, but learns directly from the data and differentiates features from the data. In terms of lung nodule detection, deep convolutional neural networks can use the training dataset to automatically select the best image features, resulting in more lung nodule features, higher accuracy, and better robustness. However, insufficient mining of data contextual information means that the existing methods still suffer from low sensitivity and a high false-positive rate.

## 2.2. Attention Mechanism

The attention mechanism refers to the fact that the human eye usually focuses more on important regions or regions of interest in an open range due to the limitation of the visual field. With the rise of convolutional neural networks, the attention mechanism has been widely used in various aspects of computer vision. This discovery has been applied to the field of Natural Language Processing (NLP), and in the literature [20], Bahdanau et al. improve translation capabilities by assigning recomputed word relevance. The most important idea is Key–Value pair (Key–Value) attention, which proposes three elements, query (Query, Q for short), key (Key, K for short), and value (Value, V for short). Through the correlation between Query and Key, the attention weight assignment to Value is achieved and the final output result is generated. This process can be briefly described as the following steps:

- (1) Input Q, K, and V elements;
- (2) The correlation/similarity between Q and K can be calculated by comparing them, a method usually measured using point multiplication, to obtain an attention score;
- (3) Normalization of the data by the softmax function to obtain the weighting coefficients;
- (4) The final value of attention is obtained by weighting and summing V according to the weighting coefficients.

Based on the traditional attention mechanism, Vaswani et al. [21] proposed a self-attention (Self-attention) mechanism, which allows the weights to be shared in the form of parameter matrices by restricting Q, K, and V to homogeneous inputs. The long-range dependency (Long-range Dependency) concept is introduced. Since the required calculations are all matrix operations, the calculation of self-attentiveness can be simplified by expressing the following equation:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

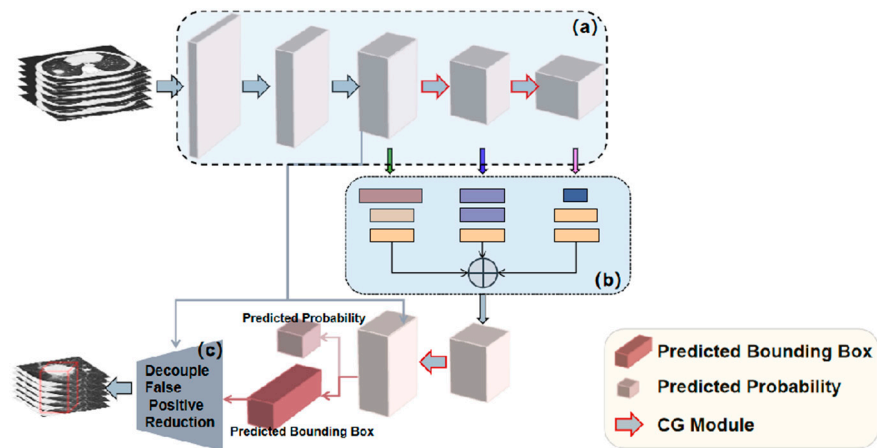
In this paper, following the characteristics of lung nodule saliency in CT images, we modify the attention mechanism to enhance the semantic representation of salient targets and strengthen the global information search capability. The module we proposed is more targeted to attention enhancement and global information search for features of lung nodule images, which will be described in detail below.

## 3. Method

Based on the observation that lung nodules can appear suddenly in lung imaging, we designed a 3DAGNet for lung nodule detection. The overall framework of 3DAGNet is shown in Figure 1, which includes an encoder backbone, 3D global search and attention enhancement components, and a multi-scale fusion unit. In particular, this work proposes a combination of global search and deep attention enhancement to target the characteristics of lung nodule prominence in CT images and integrate them into the encoding part. In this section, we will describe the mentioned method in detail.

### 3.1. Overall Framework

Our proposed 3DAGNet aims to effectively detect and identify abnormally appearing nodules in the lung to reduce the operational stress of physicians. As shown in Figure 1, in this work, a classical 3D ResNet18 is employed as the backbone, which is used to encode the slices to obtain high-level features and detailed information. The encoded and compressed feature maps are fully fused with a multi-scale cascade module. Then, the fused feature maps are up-sampled and decoded using up-sampling blocks with residual connections to the images to recover them to the appropriate size. The residual connectivity in the network can compensate for the loss of global information due to image coding, and also achieve the enhancement of advanced features and detailed information of the images. The specific structural details are shown in Table 1.



**Figure 1.** Overview of the proposed 3DAGNet. (a) Image compression path. (b) Multi-layer module. (c) False-positive reduction module. In module (b), the brown color represents the 3D convolution of size 3 and step size 2. Dark blue represents 3D deconvolution of size 2 and step size 2. Purple represents the deconvolution of size 3 with a void rate of 3. The rest of the modules are normal 3D convolution of size 3.

**Table 1.** The design and parameters of the network structure (each layer contains an activation and a 3D batch normalization).

Layer	Composition Element	Output Size
Input Feature Map		$128 \times 128 \times 128$
Priority Block	$(3 \times 3 \times 3 \text{ Conv}, S = 2, P = 1)$ $(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1)$	$64 \times 64 \times 64$
Forw 1 Block	$(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 2$ $(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 2$	$64 \times 64 \times 64$
Down Pooling 1	$2 \times 2 \times 2 \text{ MaxPooling}, S = 2$	$32 \times 32 \times 32$
Forw 2 Block	$(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 2$ $(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 2$	$32 \times 32 \times 32$
Down Pooling 2	$2 \times 2 \times 2 \text{ MaxPooling}, S = 2$	$16 \times 16 \times 16$
Forw 3 Block	$(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 2$ SACC Block $(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 3$	$16 \times 16 \times 16$
Down Pooling 3	$2 \times 2 \times 2 \text{ MaxPooling}, S = 2$	$8 \times 8 \times 8$
Forw 4 Block	$(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 3$ SACC Block $(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 3$	$8 \times 8 \times 8$
Multi-Scale Expansion		$16 \times 16 \times 16$
Back-1	$(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 2$ SACC Block $(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 2$ $2 \times 2 \times 2 \text{ DeConv}, S = 2 + [\text{Forw2}]$	$16 \times 16 \times 16$ $32 \times 32 \times 32$
Back-2	$(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 3$ $(3 \times 3 \times 3 \text{ Conv}, S = 1, P = 1) \times 3$	$32 \times 32 \times 32$

For the lung nodule model designed in this work, as shown in Figure 1, the  $X \in R^{128 \times 128 \times 128}$  is the input image of the setup, and the  $S_i = F(X)$  is the feature map output by each convolution block, where  $S_i \in R^{D \times H \times W}$ . In order to obtain more high-level feature representation and details, the acquired feature maps are down-sampled using maximum pooling after the convolution blocks. We set  $In_i = Pool(S_i)$ , where  $In_i$  represents the output result of the feature image after passing through the convolution block and pooling layers, and  $In_i \in R^{D \times H \times W}$  accordingly. Next, we input the down-sampled image into the feature enhancement and global search module, where  $S_i = Sa(F(X))$ ,  $S_i \in R^{D \times H \times W}$ , where  $Sa()$

represents the operation of the convolution block. Then, following by a series of down-sampling and feature enhancement, we take the output features of the last three layers of down-sampling as input to the multi-scale cascade module as  $S_{out} = V(S_a, S_b, S_c)$ , where  $S_i \in R^{\frac{D}{8} \times \frac{H}{8} \times \frac{W}{8}}$  represents the output feature map after the convolution block processing.  $V(\cdot)$  indicates the enhancement of image information by fusing feature information and details from multiple scales, the details of which will be explained in detail in the following subsections. Then, the image is decoded and up-sampled using a 3D deconvolution layer  $T_i = \theta(\mu(Trans(S_i)))$ ,  $T_i \in R^{D \times H \times W}$ , where  $\theta(\cdot)$  is denoted as the activation function.  $ReLU(\cdot)$  and  $\mu(\cdot)$  are represented as the batch normalization layer. Then, in order to effectively mitigate the loss due to encoding, we make up for the lost information by using residual concatenation  $T_i = Sa(cat(T_i, S_{out}))$ , where  $Cat(\cdot)$  is the summation of the two sets of feature information in the channel dimension. Using the convolution block  $Sa(\cdot)$  ensures that the feature maps are fully fused.

After the detection model output results  $T_i$ , two  $1 \times 1$  convolutions are performed to produce the category probability results  $R_{cls}$  and regression results  $R_{reg}$ . The classification results are used to calculate the loss using the binary cross-entropy function  $L_{cls}$ , with the following equation:

$$L_{bce} = -\frac{1}{N} \lim_{x \rightarrow \infty} \sum_{i=1}^N y_i \times \log(P(y_i)) + (1 - y_i) \times \log(1 - P(y_i)) \quad (2)$$

$$L_{cls} = \mu \times L_{bce}[pos] + (1 - \mu) \times L_{bce}[\varphi(neg)] \quad (3)$$

where the results are evaluated between the predicted results and the ground truth, and if the match is correct, then  $y_i = 1$  and vice versa  $y_i = 0$ .  $\mu$  represents the balance factor and  $\varphi(\cdot)$  is the process of negative sample processing using OHEM [22] algorithm according to the actual situation of negative sample. Similarly, the regression results  $R_{reg}$  use the smoothed L1 loss for error estimation,  $L_{reg} = \sum_N S(R_{reg}, GT)$ .

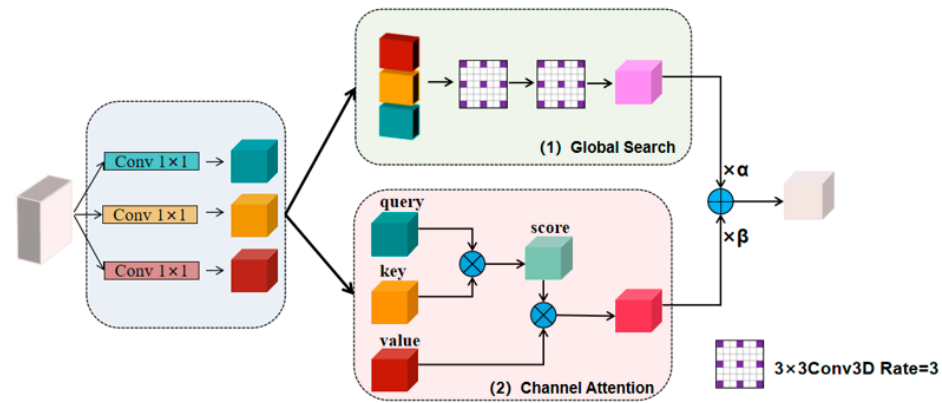
Then, to reduce the effect of false-positives on the results, the  $S_i$  and  $R_{reg}$  are fed into the false-positive reduction network, and the results are  $D = Cut(R_{reg}, S_i)$ . The loss calculation is performed on the output of the suppression network, and the cross-entropy loss is used to set the positive sample probability in the training sample as  $p$  and the predicted positive sample probability as  $q$ . The specific formula is shown below:

$$L_{cross} = -\sum_{i=1}^N (p_i \log(q_i) + (1 - p_i) \log(1 - q_i)) \quad (4)$$

The losses are finally aggregated to obtain the total of the losses of all parts for gradient back-propagation training. The following subsections explain our proposed global search and deep attention enhancement modules in detail.

### 3.2. Global and Channel (CG) Module

Following the characteristics of lung nodule saliency in CT images, the CG module aims to enhance the semantic representation of salient targets and strengthen the global information search capability. Unlike traditional feature enhancement modules, we design the module to be more targeted to attention enhancement and global information search for features of lung nodule images. Specifically, we design two branches to enhance the feature information with deep attention and global spatial information, respectively. Although the two branches enhance the feature data from different perspectives, the final output feature maps are of the same size, and then fused with an adjustable scale, and the fused feature information is input to the next stage of convolution block. The specific structure of this module is shown in Figure 2.



**Figure 2.** A total of two branches of CG module. (1) Global search branch. (2) Channel attention branch.

First, the feature maps are grouped by three  $1 \times 1 \times 1$  convolutions, and the three groups of feature data are obtained first  $G_q, G_k, G_v$ :

$$G_q = \delta(\sigma(F(S_i))) \in R^{C \times D \times H \times W} \tag{5}$$

$$G_k = \delta(\sigma(F(S_i))) \in R^{C \times D \times H \times W} \tag{6}$$

$$G_v = \delta(\sigma(F(S_i))) \in R^{C \times D \times H \times W} \tag{7}$$

where  $F(\cdot)$  represents the 3D convolutional layer of size 1, step size 1;  $\sigma(\cdot)$  represents the batch normalization layer; and  $\delta(\cdot)$  represents the  $ReLU(\cdot)$  activation function. In this work, a two-branch design is carried out in the second part for the characteristics of lung nodules in CT images, which not only enhances the image data from the perspective of deep spatial attention, but also takes into account the influence of global cues on local key information in CT images of lung nodules, and is designed to enhance the search for global information by using null convolution. The specific explanation is given below.

For the deep spatial attention enhancement branch, the first part of the given feature data  $G_q \in R^{C \times D \times H \times W}$  is transformed, through data deformation, into  $G_{q1} \in R^{C \times DHW}$  with the key elements  $G_k \in R^{C \times D \times H \times W}$  of the deformed form  $G_{k1} \in R^{C \times DHW}$  multiplied to generate the spatial attention score in the channel dimension  $G_{att} \in R^{C \times C}$ , and then multiplied with the pixel values at the same position corresponding to it to obtain  $G_{ans} \in R^{C \times DHW}$  and, finally,  $G_{ans}$  is restored to the original shape. The specific equation is as follows:

$$S_{Ci} = X + R(G_v \times \theta(R(G_q) \times R(G_k))) \tag{8}$$

where  $R(\cdot)$  denotes the reshape, and  $\theta(\cdot)$  denotes the normalization function, and the final output obtained is  $S_i \in R^{C \times D \times H \times W}$ . Then, the interpretation is unfolded by searching branches from the global information.  $G_q, G_k, G_v$  are processed separately, and all three sets of data are passed through a convolution block consisting of the dilated convolution. The specific formula is as follows:

$$S_{Di} = F_{D1}(F_{D2}(G_q, G_k, G_v)) \tag{9}$$

in which  $F_{D1}(\cdot), F_{D2}(\cdot)$  are convolutional blocks consisting of a 3D dilated convolution of size 3 and a dilated rate of 3, a batch normalization layer, and a ReLU activation layer, respectively, to enhance the global attention to the feature-informed images using the dilated convolution. Finally, the feature data obtained from the spatial attention branch

and the feature data requested from the global search branch are fused with the aim of achieving a better fusion of the two-branch feature data. The formula is shown as follows:

$$S_i = F(\alpha \times S_{Ci} + \beta \times F_{Di}) \tag{10}$$

where  $\alpha, \beta$  as the balance factor are adjustable parameters and  $F(\cdot)$  represents a 3D convolutional block of size 3.

### 3.3. Multi-Layer Module

Although the CG module has been performed to enhance the spatial and global information of the feature data, the high-level semantic information cannot be directly captured by attention enhancement alone. Therefore, we use a multi-scale cascade module to capture the multi-level information. This work takes the last three feature maps obtained from the encoding part  $S_a \in R^{C \times D \times H \times W}$ ,  $S_b \in R^{C \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}}$ ,  $S_c \in R^{C \times \frac{D}{4} \times \frac{H}{4} \times \frac{W}{4}}$  as input to the multiscale cascade module. To fuse both of multi-level semantic features and detailed information effectively, and alleviate the loss of location information, the final output is taken as the size of the intermediate feature data, i.e.,  $S_{out} \in R^{C \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}}$ . As shown in Figure 3, the details are as follows.

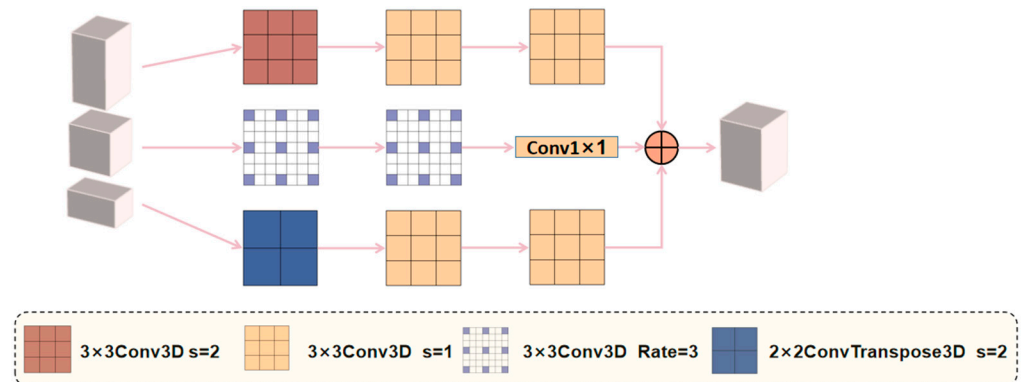


Figure 3. The designed multi-layer module.

The multiscale cascade module operates on the corresponding branch of  $S_a$  by first down-sampling the feature data using a 3D convolution of size 3 with a step size of 2 to obtain  $S_{a1}$ , where  $S_{a1} = F(S_a) \in R^{C \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}}$ . Then, the features are captured for  $S_{a1}$ . Two convolution operations are performed to obtain  $S_{a2} = RES(S_{a1})$ , where  $RES(\cdot)$  is composed of two convolution layers of size 3, a batch normalization layer and an activation function. At the same time, the middle branch  $S_b$  is performed using two dilated convolution layers of size 3 and a dilated rate of 3, and the output results are obtained  $S_{b2} = D(D(S_b))$ ,  $S_{b2} \in R^{C \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}}$ , accordingly. Similarly, an  $S_c$  up-sampling operation is performed on the corresponding branch with the following equation:

$$S_{c2} = F(F(ConvTrans(S_c))) \in R^{C \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}} \tag{11}$$

where  $F(\cdot)$  denotes a 3D convolutional layer of size 3, a batch normalization layer, and an activation function, and  $ConvTrans(\cdot)$  is a deconvolutional network layer of size 2 with a step size of 2. The results of the three branches are finally integrated by the operation of summing over the channel dimensions to obtain the final result:

$$S_{out} = F(cat(S_{a2}, S_{b2}, S_{c2})) \in R^{C \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}} \tag{12}$$

where  $cat(\cdot)$  is a stitching of the three datasets in the channel dimension and  $F(\cdot)$  is composed of two convolutional blocks with a convolutional kernel size of 3. The purpose is to achieve a better fusion of multi-scale feature data.



### 3.4. False-Positive Suppression Network

The similarity between nodule information and vascular information in lung nodule images can easily lead to more false-positive results, which can interfere with the diagnostic results to some extent. Therefore, it is particularly important to suppress false-positives. Inspired by [23], we performed a non-maximal value suppression operation for the candidate box information output from the detection network. The filtered candidate frames are then combined with the feature maps that retain more original information, and the feature maps are cropped using the candidate frame information to retain only small feature maps of  $7 \times 7 \times 7$  in the nodule region. Because of the difference in details between lung nodules and false-positive nodules, cropping can make the lung nodule information from a small target to a larger target more favorable to distinguish from details. The cropped feature maps are then input into the 3D ROI pooling module, and the ROI pooling strategy and linear neural network are used to make a more accurate feature representation of the target to achieve a more accurate classification performance and reduce the influence of false-positive nodules.

## 4. Experiment

In this section, we evaluate the performance of the designed network model on the LUNA16 nodule dataset and compare the model with several classical models in the field of lung nodule detection.

### 4.1. Datasets

We used LUNA16 to evaluate the performance of 3DAGNet. To the best of our knowledge, this dataset is composed of 888 sliced complete CT medical images extracted from the public lung nodule study dataset LIDC-LDRI, and each set of CT medical images has a slice thickness greater than 2.5 mm. For lung nodule reference, nodules greater than or equal to 3 mm identified by more than three physicians in the LUNA16 dataset are marked as reference nodules, and other conditions were not considered as reference criteria. In this study, we converted the original CT images into Housefield cells, deflated the pixel values by (0, 255), and transformed the current world coordinates corresponding to the voxel coordinates.

### 4.2. Implementation Details and Evaluation Metrics

Due to limitations in the RAM of the GPU, we cropped the images in training. Without destroying the information of the image itself, we cropped the image to a size of  $128 \times 128 \times 128$  and input it into the detection network for training. In the false-positive suppression part, based on the basic information of the nodules in the dataset, we crop the images to a size of  $7 \times 7 \times 7$  to ensure that the data are input to the second stage network without destroying the information of the nodules. Our model is built in PyTorch framework. To fit the training, we train the model with stochastic gradient descent (SGD) to 300 epochs with a batch size of 8. The initial learning rate is set at 0.01, and the decrease rate is set at 0.001 after 150 epochs, and the last 60 epochs are learned using a learning rate of 0.0001 with a momentum parameter of 0.9. We placed the entire model on a single NVIDIA RTX 3090 GPU. The six-fold cross-validation is employed for learning, and the final results are obtained jointly from the cross-validation average.

To evaluate the performance of our proposed 3DAGNet, we measure the model using the FROC (Free-response ROC) curve, which is the most commonly used objective metric for lung nodule detection. The horizontal coordinate of this metric is the false alarm rate and the vertical coordinate is the recall rate, expressed as the average sensitivity of 0.125, 0.25, 0.5, 1, 2, 4, and 8 false-positive nodules per search, which is also the official effectiveness rating metric of LUNA16.

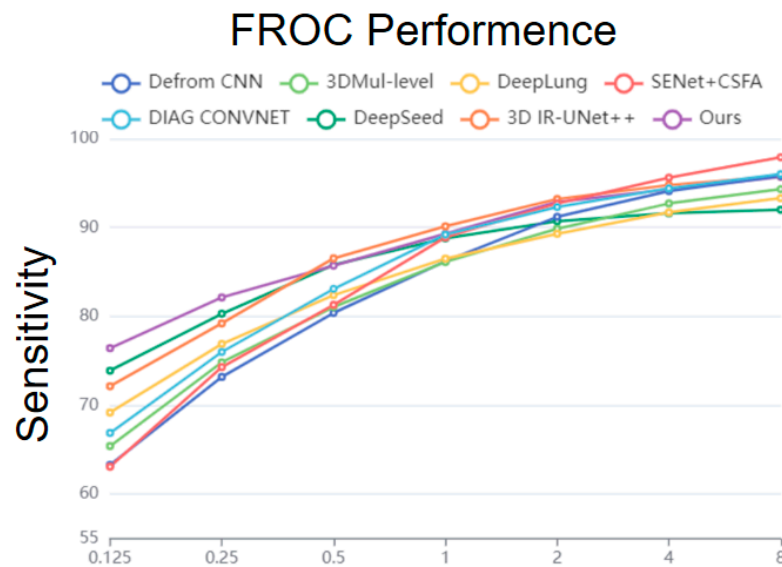
### 4.3. Comparison of Different Detection Methods

In this study, we trained and evaluated the 3DAGNet model and other classical models using the LUNA16 lung nodule dataset. The results are as follows.

Table 2 shows the results obtained based on the FROC evaluation metrics. We compared the 3DAGNet model with eight classical models in the field of lung nodule detection, containing seven classical main models: CUMedVis [24], Deform CNN [25], 3DMul-level [26], DeepLung [27], SeNet + CSFA [28], DIAG CONVNET [29], DeepSeed [23], and 3D IR-UNet++ [30]. In addition, a linear plot based on the FROC evaluation metrics (Figure 4) visually expresses that 3DAGNet has better results than other classical networks, illustrating the advanced nature of the proposed method.

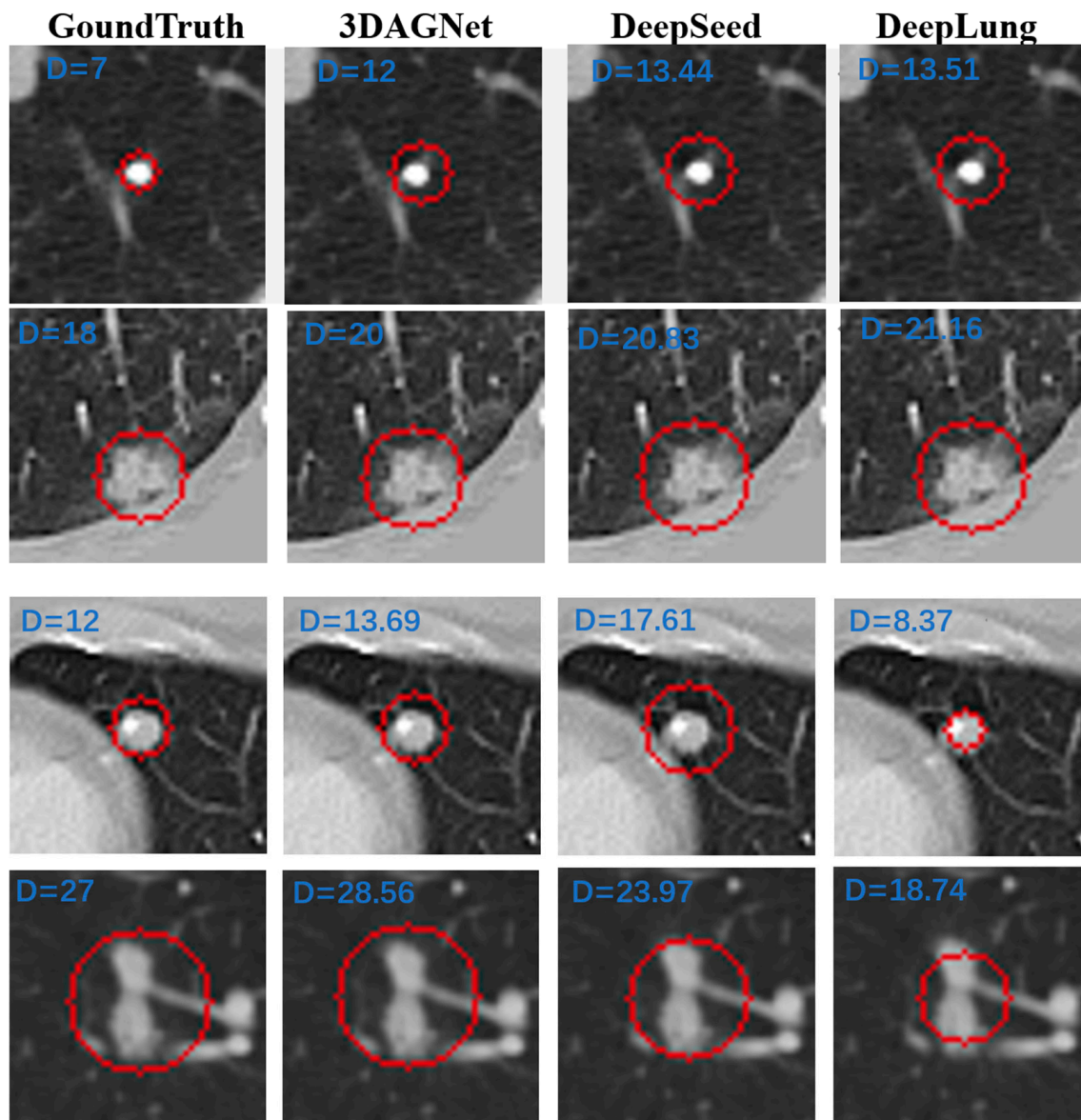
**Table 2.** Comparison of FROC evaluation metrics (%) of 3DAGNet and the other typical methods on the LUNA16 dataset.

Method	0.125	0.25	0.5	1	2	4	8	Average
CUMedVis [24]	67.70	73.70	81.50	84.80	87.90	90.70	92.20	82.63
Deform CNN [25]	63.30	73.20	80.40	86.20	91.20	94.10	95.80	83.46
3DMul-level [26]	65.43	74.84	81.06	86.13	89.86	92.70	94.31	83.47
DeepLung [27]	69.20	76.90	82.40	86.50	89.30	91.70	93.30	84.18
SENet + CSFA [28]	63.10	74.30	81.30	88.90	92.70	<b>95.60</b>	<b>97.90</b>	84.83
DIAG CONVNET [29]	66.90	76.00	83.10	89.20	92.30	94.40	96.00	85.41
DeepSeed [23]	73.90	80.30	85.80	88.80	90.70	91.60	92.00	86.16
3D IR-UNet++ [30]	72.15	79.22	<b>86.53</b>	<b>90.13</b>	<b>93.20</b>	94.77	95.78	87.39
Ours	<b>76.43</b>	<b>82.14</b>	85.71	89.29	92.86	94.29	95.71	<b>88.08</b>



**Figure 4.** The performance curve of the 3DAGNet compared with other networks based on the FROC evaluation metrics. The horizontal coordinate of this metric is the false alarm rate and the vertical coordinate is the recall rate.

In addition, in Figure 5, we visualize the test results to visually see the difference between the predicted and ground truth. In a word, the model we designed can achieve accurate predictions that infinitely approximate the true results and can reduce the diameter error between the circled target frame and the true target frame to a very low level.



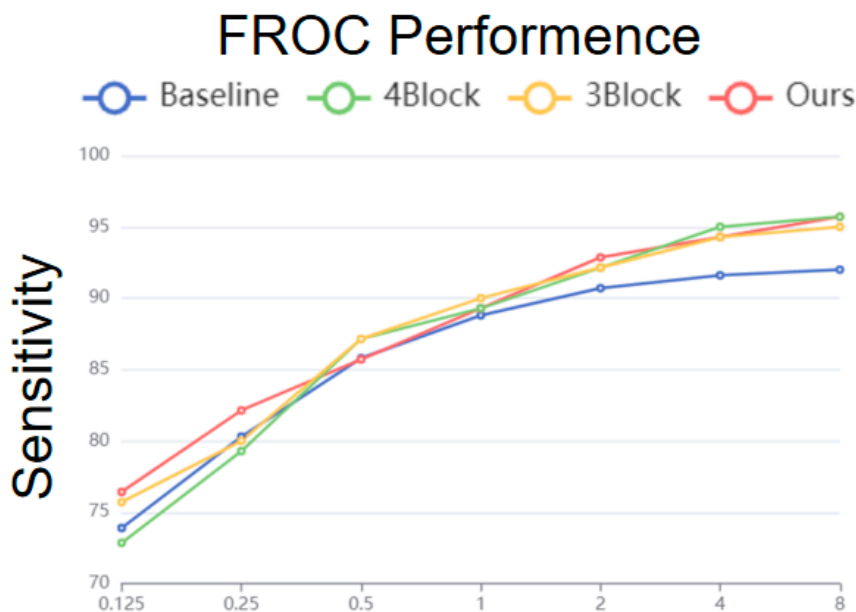
**Figure 5.** Visualization of slices of 3DAGNet model prediction results. Each envelope is represented as a defined nodule whose spatial location determines the center of the circle. The first column shows the real candidate box for the lung nodule. The second column represents the visualization information of the predicted nodule results from the 3DAGNet model. The third and fourth columns show the predicted results of the DeepSeed model and the DeepLung model, respectively.

#### 4.4. Ablation Experiment

In this study, we propose an enhancement module combining deep attention enhancement and global search, as well as a multiscale cascade module to achieve an accurate detection of lung nodules. To verify the effectiveness of the designed module and the effect of the number of modules on the network, we conducted several ablation experiments, and the comprehensive results are shown in Table 3 and Figure 6. In Figure 6, the blue FROC curve represents the 3D RPN and the FROC effect of the model with four CG modules in the encoder corresponds to the green line. Next, we ablated the modules (three CG modules were retained in the network model) and the corresponding FROC curve is shown in the orange curve below. Finally, the proposed 3DAGNet corresponds to the red FROC curve.

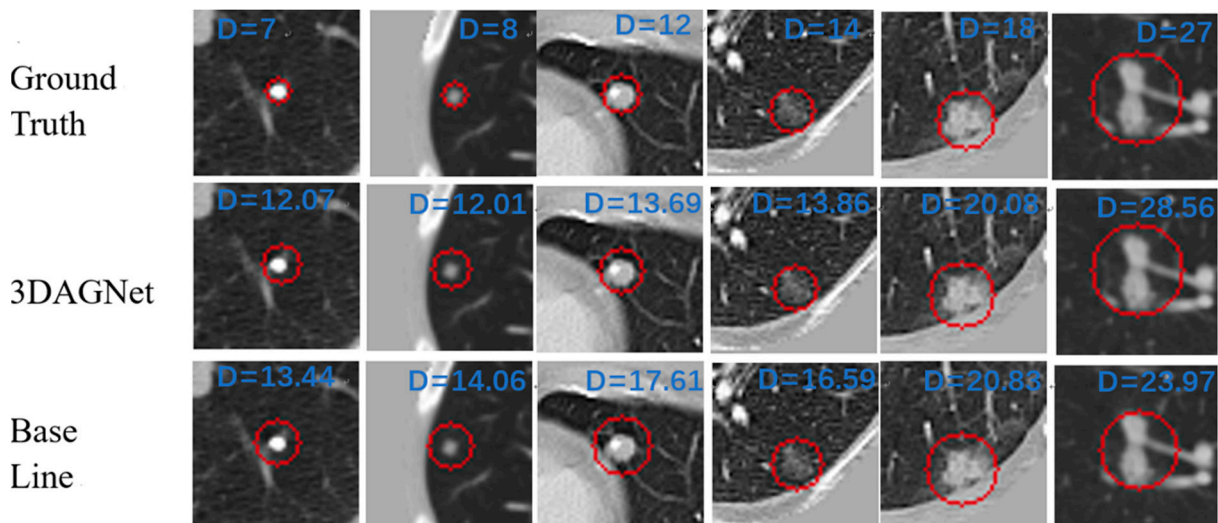
**Table 3.** Comparison of FROC evaluation metrics (%) of 3DAGNet with different attention modules implemented on the LUNA16 dataset. The baseline is 3D RPN.

Method	0.125	0.25	0.5	1	2	4	8	Average
Baseline	73.90	80.30	85.80	88.80	90.70	91.60	92.00	86.15
4 CG	72.86	79.29	87.14	89.29	92.14	<b>95.00</b>	95.71	87.35
3 CG	75.71	80.00	<b>87.14</b>	<b>90.00</b>	92.14	94.29	95.00	87.75
Ours	<b>76.43</b>	<b>82.14</b>	85.71	89.29	<b>92.86</b>	94.29	<b>95.71</b>	<b>88.06</b>



**Figure 6.** The performance curve of the 3DAGNet compared with the ablation experimental model.

Next, we visualized the results obtained from the ablation experiments. The effect of this model is visualized from a visual perspective. The details are shown in Figure 7.



**Figure 7.** Visualization of slices of the prediction results. Each circle is represented as a defined nodule whose spatial location determines the center of the circle. The first row shows the real candidate boxes for the lung nodules. The second row represents the visualization information of the predicted nodal results of the 3DAGNet model. The third row shows the predicted results of the baseline model.

The experimental results can clearly indicate that each improvement in the experimental procedure improves the detection results to some extent. 3DRPN is the least effective of the four curves, but the addition of the CG module shows a significant improvement in the performance of this network, which proves the effectiveness of the module. Through the whole ablation experiments, it can be observed that our proposed 3DAGNet has a superior effect on the detection of lung nodules.

## 5. Conclusions

In this work, we designed a novel CG module combining depth feature enhancement and global information search for the characteristics of lung nodules in CT medical images to investigate the effect of depth and global information on lung nodule detection and localization, and developed a multi-layer module integrating depth, global feature enhancement, and contextual information fusion for lung nodule detection. The results on the mainstream lung nodule detection dataset LUNA16 show that 3DAGNet is more effective than other mainstream networks for lung nodule detection. Additionally, ablation experiments demonstrated the effectiveness of the CG module and verified the usefulness of depth and global features for lung nodule detection. In conclusion, the experiments effectively proved the rationality and validity of the model structure proposed in this work.

In future work, we intend to further explore the properties of using lung nodules in CT medical imaging and design more effective modules to improve the performance of lung nodule detection.

**Author Contributions:** Methodology, M.J. and L.Z.; Software, L.Z. and H.J.; Validation, H.J. and X.L.; Formal analysis, X.L.; Data curation, X.L.; Writing—original draft, L.Z. and H.J.; Writing—review & editing, M.J. and X.L.; Supervision, M.J.; Project administration, M.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by National Science Foundation of China (NSFC) (61976123, 62072213); Taishan Young Scholars Program of Shandong Province; and Key Development Program for Basic Research of Shandong Province (ZR2020ZD44).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Conflicts of Interest:** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Szankin, M.; Kwasniewska, A. Can AI see bias in X-ray images. *Int. J. Netw. Dyn. Intell.* **2022**, *1*, 48–64. [[CrossRef](#)]
2. Wang, M.; Wang, H.; Zheng, H. A mini review of node centrality metrics in biological networks. *Int. J. Netw. Dyn. Intell.* **2022**, *1*, 99–110. [[CrossRef](#)]
3. Zhao, X.; Liu, L.; Qi, S.; Teng, Y.; Li, J.; Qian, W. Agile convolutional neural network for pulmonary nodule classification using CT images. *Int. J. Comput. Assist. Radiol. Surg.* **2018**, *13*, 585–595. [[CrossRef](#)]
4. Sori, W.J.; Feng, J.; Liu, S. Multi-path convolutional neural network for lung cancer detection. *Multidimens. Syst. Signal Process.* **2019**, *30*, 1749–1768. [[CrossRef](#)]
5. Tang, H.; Kim, D.R.; Xie, X. Automated pulmonary nodule detection using 3D deep convolutional neural networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 523–526.
6. Yao, F.; Ding, Y.; Hong, S.; Yang, S.H. A survey on evolved LoRa-based communication technologies for emerging internet of things applications. *Int. J. Netw. Dyn. Intell.* **2022**, *1*, 4–19. [[CrossRef](#)]
7. Yu, N.; Yang, R.; Huang, M. Deep common spatial pattern based motor imagery classification with improved objective function. *Int. J. Netw. Dyn. Intell.* **2022**, *1*, 73–84. [[CrossRef](#)]
8. Murphy, K.; van Ginneken, B.; Schilham, A.M.R.; De Hoop, B.J.; Gietema, H.A.; Prokop, M. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Med. Image Anal.* **2009**, *13*, 757–770. [[CrossRef](#)] [[PubMed](#)]

9. Lee, S.L.A.; Kouzani, A.Z.; Hu, E.J. Random forest based lung nodule classification aided by clustering. *Comput. Med. Imaging Graph.* **2010**, *34*, 535–542. [[CrossRef](#)] [[PubMed](#)]
10. Sui, Y.; Wei, Y.; Zhao, D. Computer-aided lung nodule recognition by SVM classifier based on combination of random undersampling and SMOTE. *Comput. Math. Methods Med.* **2015**, *2015*, 368674. [[CrossRef](#)]
11. Chen, C.; Xie, Y.; Lin, S.; Yao, A.; Jiang, G.; Zhang, W.; Qu, Y.; Qiao, R.; Ren, B.; Ma, L. Comprehensive regularization in a bi-directional predictive network for video anomaly detection. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 230–238. [[CrossRef](#)]
12. Lin, S.; Ji, R.; Li, Y.; Deng, C.; Li, X. Toward Compact ConvNet via Structure-sparsity Regularized Filter Pruning. *IEEE Trans. Neural Netw. Learn. Syst. (TNNLS)* **2020**, *31*, 574–588. [[CrossRef](#)] [[PubMed](#)]
13. Li, Y.; Lin, S.; Liu, J.; Ye, Q.; Wang, M.; Chao, F.; Yang, F.; Ma, J.; Tian, Q.; Ji, R. Towards compact cnns via collaborative compression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 19–25 June 2021; pp. 6438–6447.
14. Sun, R.; Pang, Y.; Li, W. Efficient Lung Cancer Image Classification and Segmentation Algorithm Based on an Improved Swin Transformer. *Electronics* **2023**, *12*, 1024. [[CrossRef](#)]
15. Li, W.; Cao, P.; Zhao, D.; Wang, J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Comput. Math. Methods Med.* **2016**, *2016*, 6215085. [[CrossRef](#)] [[PubMed](#)]
16. Zhu, X.; Wang, X.; Shi, Y.; Ren, S.; Wang, W. Channel-Wise Attention Mechanism in the 3D Convolutional Network for Lung Nodule Detection. *Electronics* **2022**, *11*, 1600. [[CrossRef](#)]
17. Jin, H.; Li, Z.; Tong, R.; Lin, L. A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. *Med. Phys.* **2018**, *45*, 2097–2107. [[CrossRef](#)]
18. Mei, J.; Cheng, M.M.; Xu, G.; Wan, L.-R.; Zhang, H. SANet: A slice-aware network for pulmonary nodule detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4374–4387. [[CrossRef](#)]
19. Song, T.; Chen, J.; Luo, X.; Huang, Y.; Liu, X.; Huang, N.; Chen, Y.; Ye, Z.; Sheng, H.; Zhang, S. CPM-Net: A 3D center-points matching network for pulmonary nodule detection in CT scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2020; pp. 550–559.
20. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
21. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inform. Process. Syst.* **2017**, *30*. [[CrossRef](#)]
22. Shrivastava, A.; Gupta, A.; Girshick, R. Training Region-based Object Detectors with Online Hard Example Mining. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vegas, NV, USA, 27–30 June 2016; pp. 27–30.
23. Li, Y.; Fan, Y. DeepSEED: 3D squeeze-and-excitation encoder-decoder convolutional neural networks for pulmonary nodule detection. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1866–1869.
24. 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*, 1558–1567. [[CrossRef](#)]
25. Yuan, H.; Fan, Z.; Ding, D.; Sun, Z. False-positive reduction of pulmonary nodule detection based on deformable convolutional neural networks. In Proceedings of the 2021 IEEE 9th International Conference on Bioinformatics and Computational Biology (ICBCB), Taiyuan, China, 25–27 May 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 130–134.
26. Lu, X.; Gu, Y.; Yang, L.; Zhang, B.; Zhao, Y.; Yu, D.; Zhao, J.; Gao, L.; Zhou, T.; Liu, Y.; et al. Multi-level 3D densenets for false-positive reduction in lung nodule detection based on chest computed tomography. *Curr. Med. Imaging* **2020**, *16*, 1004–1021. [[CrossRef](#)]
27. Zhu, W.; Liu, C.; Fan, W.; Xie, X. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 673–681.
28. Gu, Z.; Li, Y.; Luo, H.; Zhang, C.; Du, H. Cross attention guided multi-scale feature fusion for false-positive reduction in pulmonary nodule detection. *Comput. Biol. Med.* **2022**, *151*, 106302. [[CrossRef](#)] [[PubMed](#)]
29. Setio, A.A.A.; Ciompi, F.; Litjens, G.; Gerke, P.; Jacobs, C.; van Riel, S.J.; Wille, M.M.; Naqibullah, M.; Sanchez, C.I.; van Ginneken, B. Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* **2016**, *35*, 1160–1169. [[CrossRef](#)] [[PubMed](#)]
30. Lin, J.; She, Q.; Chen, Y. Pulmonary nodule detection based on IR-UNet++. *Med. Biol. Eng. Comput.* **2023**, *61*, 485–495. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.