Original Article

A lightweight neural network for lung nodule detection based on improved ghost module

Liuyang Yang[1#], Hongyu Cai[1#], Xinyu Luo[1#], Jianping Wu[2#], Rui Tang[1], Yu Chen[1], Wei Li[3]

[1]Department of Management Science and Information System, Faculty of Management and Economics, Kunming University of Science and Technology, Kunming, China; [2]Department of Radiology, Yunnan Cancer Hospital & the third Affiliated Hospital of Kunming Medical University, Kunming, China; [3]The First People's Hospital of Yunnan Province, Kunming, China

Contributions: (I) Conception and design: R Tang; (II) Administrative support: R Tang, Y Chen, W Li; (III) Provision of study materials or patients: R Tang, J Wu; (IV) Collection and assembly of data: L Yang, X Luo; (V) Data analysis and interpretation: X Luo, H Cai; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work and should be considered as co-first authors.

Correspondence to: Rui Tang, PhD. Department of Management Science and Information System, Kunming University of Science and Technology, 68 Wenchang Road, 121 Street, Kunming 650093, China. Email: tangrui@kust.edu.cn.

Background: Computer tomography images are the preferred method of preoperative evaluation for lung disease. However, it remains difficult to detect and recognize nodules accurately and efficiently due to poor data imaging quality, heavy reliance on physician experience and the need for more human-computer interaction for diagnosis. Currently, image nodule detection based on deep convolutional neural networks has gained much momentum.

Methods: To alleviate doctors' tremendous labor in the diagnosis procedure, and improve the accuracy of intelligent detection of lung nodules, we improved GhostNet and proposed a lightweight neural network for object detection for lung nodule image detection. Firstly, the bneck structure in the backbone feature extraction network is adopted and improved from the structure of MobileNetV3. The weights are adjusted by changing the initial channel attention mechanism and introducing a spatial-temporal attention mechanism. Then, in the enhanced feature extraction part, we mainly use depth-separable convolution blocks to replace the 3×3 convolution of the original network for the purpose of reducing the model parameters, and make more improvements based on the network structure to enhance the applicability of the network. Diagnostic precision, recall, F1-score, mAP and parameter count were calculated.

Results: According to our lightweight neural network, F1-score, precision, and recall were 0.87, 86.34%, and 86.69%, respectively. Based on our dataset, the Yolov4-GNet network proposed in this research outperforms the current neural networks on both precision and recall as well as F1.

Conclusions: The lung nodule detection method proposed in this research not only simplifies the processing of images, but also outperforms comparable methods in nodule detection rate and positioning accuracy, providing a new way for lung nodule detection.

Keywords: Lung nodules; lightweight neural network; object detection; deep learning

Submitted Dec 06, 2021. Accepted for publication Mar 29, 2023. Published online Apr 17, 2023.
doi: 10.21037/qims-21-1182
View this article at: https://dx.doi.org/10.21037/qims-21-1182

Introduction

Lung cancer, as one of the most common and most prevalent diseases, has become a serious threat to people's health worldwide (1,2). There are many causes of lung cancer, including smoking, environmental and occupational factors, ionizing radiation, and chronic lung diseases, which not only lead to a continuous increase in the number of lung cancer patients, but also put new demands on the diagnosis of lung cancer. The detection of lung cancer begins with the diagnosis of a lung nodule, the diameter size of which is an important indicator of its benignity or malignancy. Lung nodules are nodular shadows ≤30 mm in diameter, usually small, located in the lung parenchyma, and lacking clinically specific symptoms and signs. Especially in computed tomography (CT) images, there are usually only a small number of pixels, which are difficult to detect and susceptible to noise interference, resulting in inaccurate localization of tissue regions. On the other hand, the lung structure is complex, and the pathological features of lung nodules are diverse. Lung nodules attached to normal tissues such as lung walls and blood vessels are more closely associated with surrounding tissues and are not easily detected, and atypical lung nodules are not easily identified, which poses a challenge to the early treatment of lung cancer. Early detection of lung nodules can significantly improve the prognosis of lung cancer and reduce mortality (3). However, previous manual methods of detection and analysis are inadequate for the massive amount of image data generated. The diagnostic process mainly relies on physicians' operational methods and practical experience, imposing a significant burden on their work.

As medical imaging technology continues to develop, traditional detection methods are struggling to cope with the increasing size and complexity of datasets. Manual feature extraction processes are also complicated and fail to effectively mine the abundant information contained in images. Recently, the application of artificial intelligence in medicine has addressed several complex medical challenges, and image detection has emerged as a new research focus (4-6). In image classification, AI techniques identify and differentiate images based on different features extracted from a large number of images for the purpose of intelligent diagnosis. At present, deep learning for lung nodule detection has also become a hot research topic (7,8). Traditionally, diagnostic classification models were established using image classification technology that manually extracted and screened features (9). As the amount of data gradually increases and the required features become increasingly complex, the manual screening of features can no longer satisfy the application demand. Deep learning generates efficient detection models through iterative training, learning and feedback using known datasets to enable automatic extraction and filtering of classification features. Despite unprecedented results in image classification and target detection, deep learning still presents many difficulties in medical imaging.

In this research, we improved the state-of-the-art object detection network model for intelligent automatic segmentation of lung nodules by developing a full convolutional neural network (CNN) with a novel feature pyramidal attention mechanism. For large scale dependencies, we use spatial-temporal attention to guide the model to focus more on global features in images. The spatial-temporal module computes pixel similarities, which can help our network identify spatial structure. Meanwhile, the model parameters and calculation consumption are significantly reduced by using depthwise separable convolution. Furthermore, we have simplified the framework of our proposed model to make it more computationally efficient and easier to train. The proposed lightweight CNN model can work well on large datasets but also scale well to small datasets without overfitting. Specifically, the contributions of this research can be summarized as follows: (I) the latest object detection model was improved and applied to lung nodule detection. The improved model dynamically incorporates local to global lung nodule features, thus reducing the drawback of incomplete extracted information caused by manually designed features. (II) The deep learning-based object detection algorithm has good generality and generalization ability, so it is easy to detect some atypical nodules. It improves the problem that the object detection algorithm is difficult to identify small targets and has poor accuracy, and effectively improves the accuracy of computer-aided lung nodule detection. (III) We introduce a spatial-temporal attention mechanism in the backbone network to maintain a large range of dependencies, allowing our network to find cues near the lung nodule region, which are crucial for capturing the distinguished regions used for feature extraction in CNN.

Methods

The study was conducted in accordance with the

Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of the Yunnan Cancer Hospital. The requirement for written informed consent was waived due to the retrospective nature of the study.

Accurate detection of lung nodules is a crucial step in the diagnosis of lung cancer, as they represent an early manifestation of the disease. In the early stages of lung nodule detection, researchers mainly used traditional machine learning methods and feature engineering methods for lung nodule detection (10,11). With the development of medical imaging technology, traditional detection methods are difficult to adapt to the current large and complex data sets, while the steps of manual extraction of features are complicated and cannot effectively mine the rich information in images. The neural network deep learning method is based on a powerful feature recognition function that can, by learning and analyzing a large amount of data, automatically find and extract regular features to achieve superior classification and diagnostic results (12-14). Recent deep learning-based approaches, especially CNNs, automatically learn powerful 3D features in an end-to-end manner with promising generalization performance, primarily for detection and classification in the diagnosis of lung nodules (15). Deep learning generates efficient classification models by repeated training, learning and feedback using known datasets for automatic extraction and filtering of classification features. The application of deep learning models for lung nodule diagnosis plays an important role in early detection of lung nodules, improving the survival rate and prolonging the survival time of lung cancer patients, and reducing the rate of misdiagnosis by physicians.

YOLO (16), You Only Look Once, is a classical one-stage object detection algorithm with fast computational speed and simple structure of the whole model. It has the advantages of efficiency, flexibility and good generalization performance. Compared with the two-stage object detection algorithm, YOLO uses a neural network for end-to-end training and then directly predicts the location and class of the object. Although it is not as accurate as the two-stage model, it has significant advantages in terms of detection speed and timeliness. YOLOv4, as a mainstream object detection model, has high detection accuracy and detection speed. In this research, it is used as a base model for lung nodule detection, and a series of improvements are made on this model so that the model can be applied to the object detection task under a specific category and achieve better detection results.

Data acquisition

Data acquisition involves obtaining three-dimensional images of a patient's lungs through various imaging techniques. Biomedical images, such as computed tomography, chest x-ray, and magnetic resonance imaging (MRI), are commonly used for detecting lung nodules. Among these, lung computed tomography requires 3D images taken from multiple angles to sample lung information, and the images are combined into multiple scans. Compared to other image scanning techniques, the higher definition of the test makes it easier to identify microscopic lung nodules, thus helping doctors to accurately detect and treat them in a timely manner. The accuracy of deep learning models heavily relies on the quality of the training dataset and labeling. The weight parameters of the deep CNN are determined by multiple iterations of training with the acquired data. Therefore, obtaining a reliable and well-labeled medical image dataset for lung cancer is a prerequisite for model training and testing. For this purpose, we invited experts from a tertiary care hospital in Yunnan Province to process and annotate some of the lung imaging data in their hospital, while anonymizing the data processing in order to protect patient privacy. The lung nodules are marked with bounding boxes and the assigned benign/malignant class labels. Finally, we obtained a total of 254 valid datasets.

The availability of high-quality labels for the detected targets in the images directly affects the training performance of the model. Incorrect or imprecise labeling can negatively affect the program's ability to recognize targets, ultimately compromising the effectiveness of model training. To annotate the location, size, and shape of lung nodules in our extracted dataset, physicians were tasked with manual labeling due to the complexity of medical images. In this research, we employed the LabelMe labeling software to improve the accuracy of our labeling. *Figure 1* shows how we labeled the lung nodules with high quality, and the CT scan of a patient with a lung nodule in the upper left lung (the right half of the patient's left lung appears in the CT image). *Figure 1*[1-6] represents the order of the images. In each image, the left half is the unlabeled image and the right half is the labeled image. For better viewing, the top half is the original size image and the bottom half is the partially enlarged image. The size and shape of this patient's lung nodules are different in the different pictures because the CT machine takes intermittent pictures while moving. In order to label the lung nodules more accurately, in each different

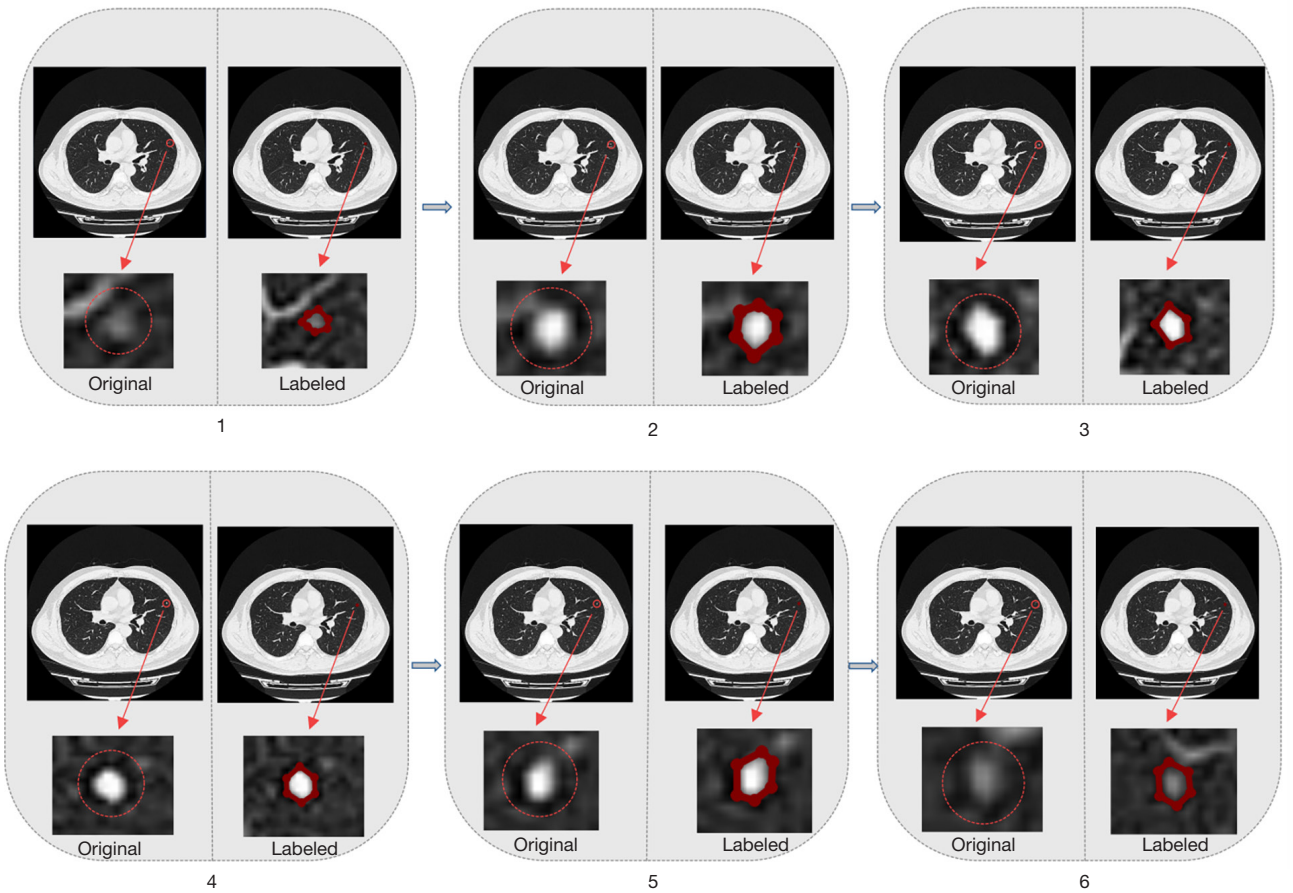

Figure 1 Labeling process on lung nodule images.

picture, we have drawn polygons of different shapes to label the lung nodules according to their morphology.

To demonstrate the detailed process of identifying lung nodules, *Figure 2* shows a magnified image of the lung and illustrates the different sizes of lung nodules in the dataset; *Figure 2*[1] shows a nodule measuring approximately 12 mm × 8 mm, *Figure 2*[2] shows a nodule measuring approximately 8 mm × 5 mm, and *Figure 2*[3] shows a nodule measuring approximately 3 mm × 2 mm. These dimensions were measured by a medical professional on the CT image review software to arrive at a diagnosis. To address the question of how to determine lung nodules, a dual validation method was used in this research. When the original images were obtained from the hospital, the diagnostic description in the form of text was already included next to the images, and the images and the diagnostic information were saved separately. Then two doctors were invited to perform the annotation independently, and the final annotation results were saved. Both the original diagnosing physicians and the two invited physicians are experts who graduated from the imaging medicine program and have more than 5 years of experience in image identification and diagnosis of lung nodules. A dual validation method is used to ensure that the lung nodules are correctly and accurately identified.

Spatial-temporal attention

CT images of the lungs are typically imaged dozens to hundreds at a time, and all CT images constitute dynamic CT findings of the patient from all angles over a short period of time. When the CT scanner is moving in the patient's lungs, the lung nodules may be obscured due to the different angles, thus causing misdiagnosis medical errors. Therefore, a spatial-temporal attention mechanism is proposed to track the lung nodules as much as possible

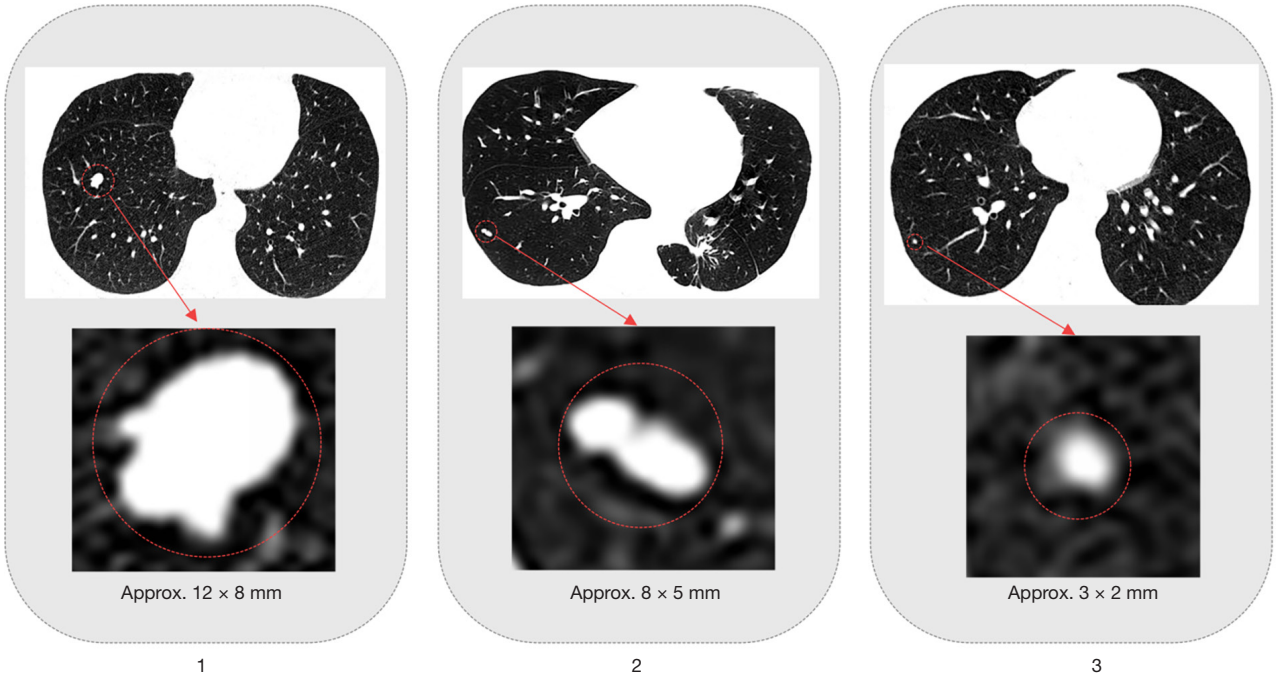

Figure 2 Lung nodules of different sizes. 1: 12 mm × 8 mm, 2: 8 mm × 5 mm, 3: 3 mm × 2 mm.

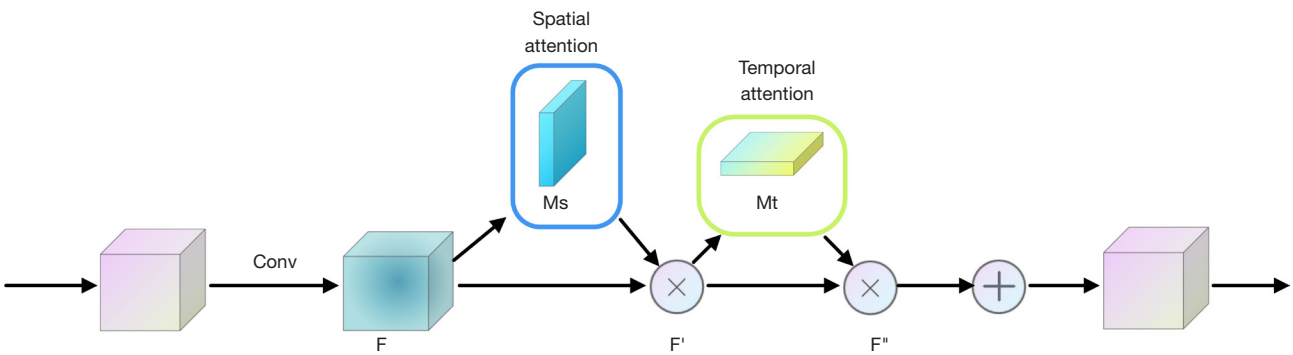

Figure 3 The spatial-temporal attention mechanism.

by analogy with the convolutional attention mechanism. As shown in *Figure 3*. The spatial-temporal attention mechanism is a combination of spatial attention mechanism and temporal attention mechanism together (17), which captures the correlation of images and fusion with the input features, respectively. When the images are input to the network, they first go through the spatial attention module to extract the spatial location feature map, and then each target is predicted on this spatial location feature map. For the extracted spatial information, if a suspected lung nodule is confirmed to exist at that location, a CNN branch is assigned to it for tracking, i.e., as many lung nodules as there are CNN branches. And the temporal attention response considers the weight of this frame prediction tracking frame accounting for the loss when we update the parameters. The spatial attention mechanism captures the dynamic spatial correlation between different locations, while the temporal attention mechanism captures the dynamic temporal correlation between different times. In particular, the output features of the spatial-temporal attention module have the same dimension as the input, and the module can be easily embedded between any

convolutional layers. By introducing a spatial-temporal attention mechanism to adjust the backbone network feature extraction weights, aiming to help the model focus on the most discriminative information and attenuate the influence of redundant information on the recognition effect.

Loss function

The common loss optimizations are Intersection over Union (IOU), Generalized IOU (GIOU), Distance IOU (DIOU), Complete IOU (CIOU), etc. Among them, IOU is the most widely used algorithm, which is defined by the following equation.

$$IOU = \frac{TP}{FP + TP + FN} \quad [1]$$

where TP, FP and FN indicate true positive, false positive and false negative counts, respectively.

When the detection box and the real box do not overlap, the IOU loss is the same, and GIOU adds the smallest rectangular box containing the detection box and the real box, so that it can solve the problem that the detection box and the real box do not overlap. But when the phenomenon of inclusion between the detection box and the real box appears GIOU and IOU loss is the same effect. DIOU takes into account the shortcomings of GIOU and includes both the real frame and the predicted frame on the basis of GIOU, but DIOU calculates the Euclidean distance between each detection frame instead of the intersection between frames, so that the problem of GIOU inclusion can be solved.

The CIOU adds the loss of detection frame scale, as well as the loss of length and width on the basis of DIOU, so that the prediction frame is more consistent with the real frame. Considering the target-anchor distance, overlap rate, scale and penalty terms, CIOU makes the target frame regression more stable, and will not have problems like IOU and GIOU such as scattering during training. And the penalty factor takes into account the aspect ratio of the predicted frame to fit the target frame. The CIOU formula is as follows:

$$CIOU = IOU - \frac{\rho^2\left(b, b^{gt}\right)}{c^2} - \alpha v \quad [2]$$

Where $\rho^2\left(b, b^{gt}\right)$ represents the Euclidean distance between the centroids of the prediction box and the real box, respectively. The c represents the diagonal distance of the smallest closed region that can contain both the prediction frame and the real frame.

$$\alpha = \frac{v}{1 - IOU + v} \quad [3]$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \quad [4]$$

The loss expression of CIOU is as follows.

$$LOSS_{CIOU} = 1 - IOU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \quad [5]$$

In addition, when the object detection platform classifies pixel points using softmax, it uses Cross Entropy Loss, which is represented by the following equation.

$$Loss = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic} \log\left(p_{ic}\right) \quad [6]$$

where M represents the number of categories classified; y_{ic} is the sign function (0/1), which takes 1 if the true category of sample i is equal to c and 0 otherwise; y_{ic} is the predicted probability that the observed sample i belongs to category c.

Network architecture

The network is built on the basis of YOLOv4 with some improvements to YOLOv4. Specifically, the network is divided into G-BNet, SPP, PANNet and Head sections. First, there is the backbone network part. In this paper, we borrow the structure of GhostNet (18), which mainly consists of the Ghost module (*Figure 4*). The main convolution in the Ghost module can have a custom kernel size, and first generate some intrinsic feature mappings with ordinary convolution, and then augment the features with ordinary linear operations to add channels. The linear operations in this module can have a great diversity. In addition, the feature mappings can be parallel to the linear transformations in the Ghost module to preserve the intrinsic feature mappings. The Ghost module with a special structure can produce many re-imaged feature mappings from inexpensive operations that can fully reveal the information behind the intrinsic features. The Ghost bottleneck structure can be constructed on top of the Ghost module, as shown in *Figure 5*. The structure is divided into two B-Neck structures, stride =1, and stride =2, where the B-Neck with step size 2 has one more step of depthwise convolution structure than the B-Neck with step

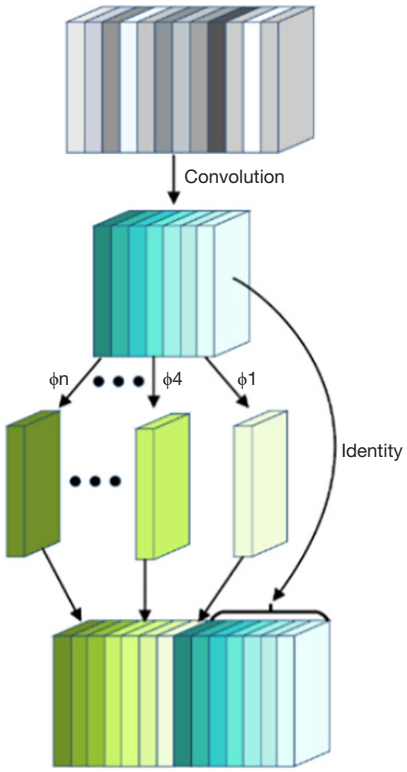

Figure 4 Ghost module.

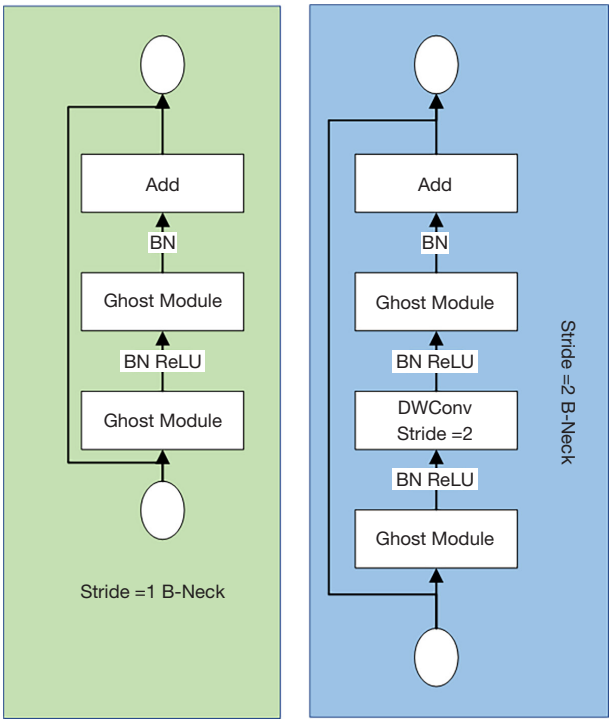

Figure 5 Original Ghost bottleneck. BN, batch normalization; DWConv, depthwise convolution.

size 1, and the input side is occupied by 1×1 convolution, called pointwise convolution. The input data is unfolded by a layer of Ghost module, and then passed into Ghost module by batch normalization and ReLU function to achieve the reduction of the number of channels to the same number of channels as the original input, and then the data processed in this step is normalized again by batch normalization. After this step, it can be stitched with the original data by residual edges to achieve the purpose of feature extraction. The overall efficiency is simple. Finally, the GhostNet structure is formed by overlaying the Ghost bottleneck structure. In view of the excellent performance of Ghost bottleneck, this study uses Ghost module and Ghost bottleneck structure in improving the network. To improve the feature extraction performance, we construct the improved Ghost bottleneck, referred to as G-Bneck, by adding the spatial-temporal attention mechanism to the original Ghost bottleneck. Specifically, the spatial-temporal attention mechanism is added to the second Gohst module of the Ghost bottleneck, and then batch normalization is performed. Finally, the inputs and outputs of the two ghost modules are connected using residual edges. The improved G-Bneck is shown in *Figure 6*.

Incidentally, the depthwise and pointwise convolutions used in G-BNeck can be combined into a depth-separable convolution structure (*Figure 7*). Depthwise separable convolution works in a similar way to standard convolution, except that it divides standard convolution into two steps. Depthwise convolution is first performed independently for each channel of the input, and one convolution kernel of depthwise convolution is responsible for one channel, while a channel is convolved by only one convolution kernel. This process produces a feature map with exactly the same number of channels as the input ones. But depthwise convolution does not extend the feature map, because the number of feature maps is the same as the number of channels in the input layer. Moreover, this operation convolves each channel in the input layer independently and cannot effectively utilize the feature information of different channels at the same spatial location.

Therefore, pointwise convolution is needed to combine these feature maps. The pointwise convolution operation is very similar to the traditional convolution operation with a convolution kernel of size $1 \times 1 \times M$, where M is the number of channels in the previous layer. So, the

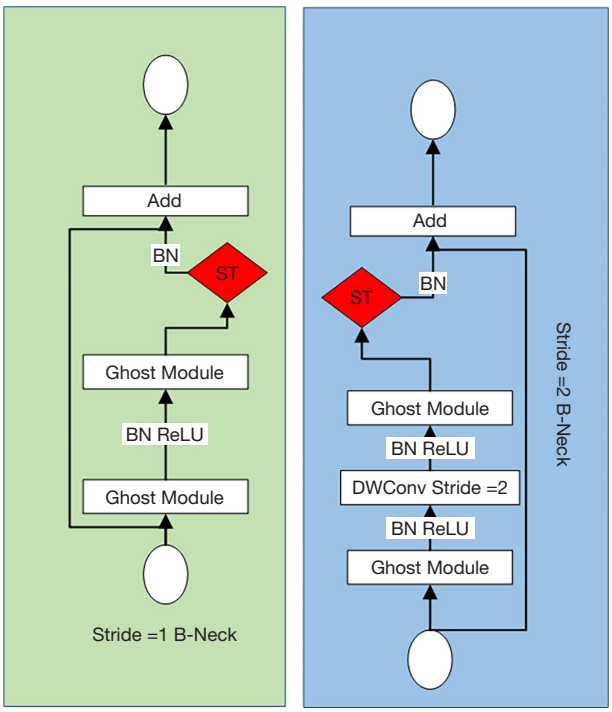

Figure 6 Improved Ghost bottleneck (G-BNeck). BN, batch normalization; DWConv, depthwise convolution; ST, spatial-temporal attention mechanism.

convolution operation here combines the feature maps in the depth direction from the previous step to generate a new feature map. Compared with standard convolution, depth-divisible convolution effectively reduces the number of parameters in the convolution operation. The ratio of computational consumption is only related to the number and size of convolution kernels. In this model, all the places where ordinary convolution is required use this structure, and both the number of parameters and the computational consumption are greatly reduced.

Spatial pyramid pooling (SPP) is a special kind of pooling layer that is incorporated into the construction process of this network algorithm because of the simple operation principle and significant effect of this structure. After the image is input, it first passes through the convolutional layer, and then three maximum pooling operations are performed with convolutional kernels of different sizes for multi-scale feature extraction, and then the information from the three pooling is stitched into a size acceptable to the fully connected layer. The principle of SPP is shown in *Figure 8*.

In terms of feature fusion and output, the YOLOv4 approach is retained. As shown in *Figure 9*, in the PANNet part, feature fusion is first performed by up-sampling, and then feature transfer is performed bottom-up using down-sampling to achieve parameter fusion of different layers of the backbone network. Finally, fusion is performed by the head section for prediction output.

Based on these preparations, a lightweight neural network is proposed as a method to detect nodules images from lung nodule images. The model enhances the learning ability of CNN and can be lightweight while maintaining accuracy. They can drastically reduce the number of parameters and computational cost. Specifically, when images are fed into the network, they are first convolved once, BN normalized once, and then the h-swish activation function is added, followed by connecting six G-Bneck modules as described in the previous section. The first five of these G-Bneck modules apply stride =1, and the sixth G-Bneck module applies stride (2) for feature extraction. A branch is left at this node for convolution and then (noted as A1) awaits subsequent operations. The backbone continues to connect 6 G-Bneck networks with the same operation as before, applying stride =1 to the first 5 G-Bneck modules and stride =2 to the sixth G-Bneck module for feature extraction. At the end of the operation, one branch is still branched out and awaits subsequent operations after convolution (noted as A2). The backbone network continues to connect 3 G-Bneck structures, where the first 2 G-Bneck modules apply stride =1 and the 3rd G-Bneck module applies stride =2 for feature extraction. At this point, the backbone network G-BNet is constructed. The images arrive from the input where the initial feature extraction work has been performed, then after picking up a 3×3 convolution, they are fed into the SPP network structure to change the size and increase the field of view to prevent feature misses. Then another 3×3 convolution layer is passed, and the feature map here is convolved (noted as A3) and split into two routes, where the feature map on one of the routes undergoes a convolution and upsampling operation and then is spliced with the feature map A2, which was branched out for the second time in the previous section. After splicing, 5×5 convolution (A32) is performed, and it continues to be divided into two ways, one of which is left untouched while the other continues to convolve and upsample once. After the up-sampling is completed, it is spliced with the feature map A1 of the first branch. And after the splicing is completed, it is divided into two branches (noted as A321) after 5×5 convolution. One of them is sent to Head1 of the head network for one

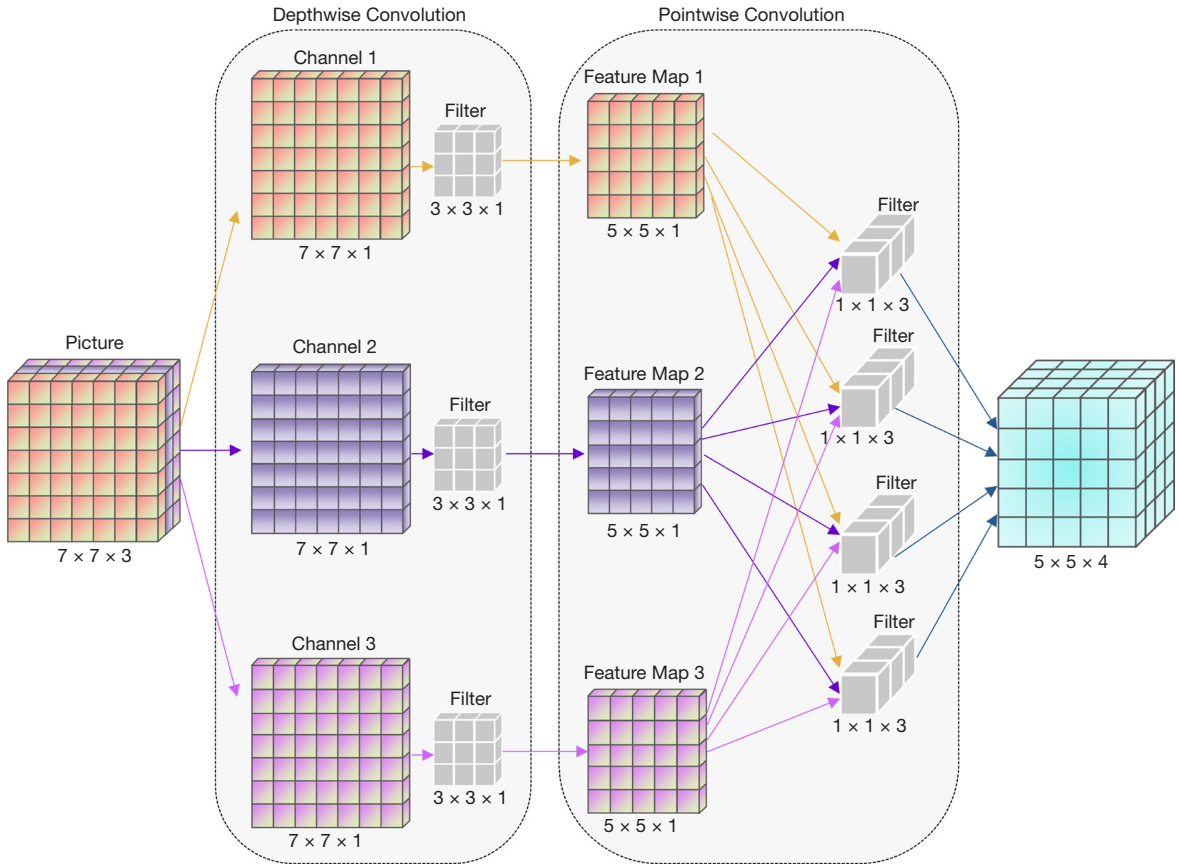

Figure 7 Depthwise separable convolution.

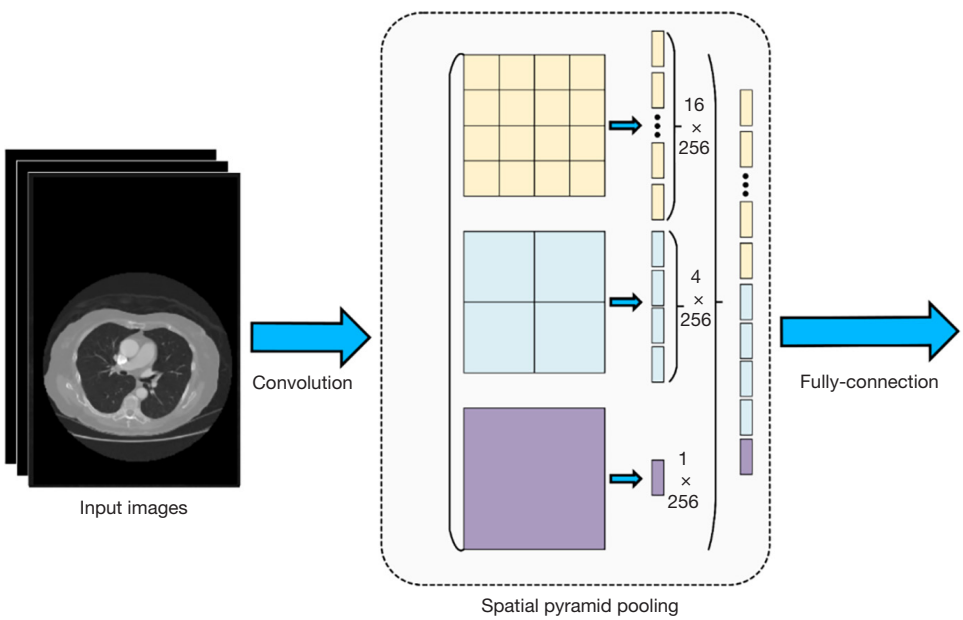

Figure 8 The principle of spatial pyramid pooling.

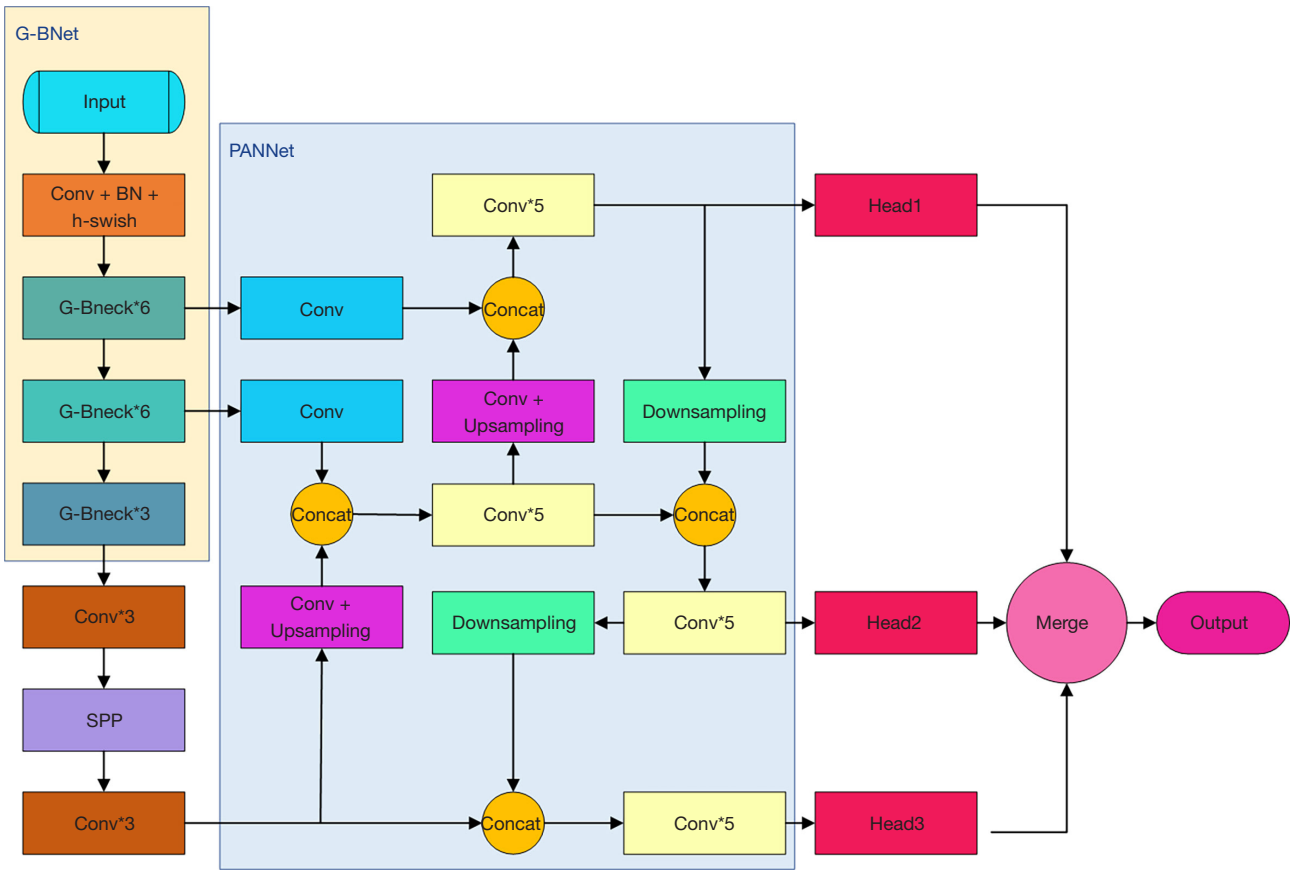

Figure 9 Network structure. Conv, convolution; BN, batch normalization; G-Bneck, improved Ghost bottleneck; Conv, convolution; SPP, spatial pyramid pooling.

feature prediction, and the other one is down-sampled and spliced with the feature map A32 reserved in the previous section, and after splicing, it undergoes 5×5 convolution (noted as A132). The convolved feature map continues to be divided into two, one of which is sent directly to the head network Head2 for one prediction, and the other is downsampled once. The other one is down-sampled and spliced with the original retained feature map A3, which is then 5×5 convolved and sent to the head network Head3 for one prediction. Finally, the three prediction structures are fused and the prediction results are output through a fully connected layer.

Results

In this section, we first describe the experimental procedure of the model and the performance results of processing the lung nodule dataset, and perform several experiments to demonstrate the effectiveness of the spatial-temporal attention module. After that, we report the experimental results of several detection methods on the lung nodule dataset, which will demonstrate the accuracy and efficiency of our detection methods.

Data enhancement

Deep CNNs require a large amount of data to train the model due to the large number of parameters present in the network. Increasing the amount of data can improve the training of the model and also prevent the overfitting of the model due to the small amount of data. Data augmentation can be achieved by adding new data or enhancing existing data. However, since the cost of acquiring new data is higher than the cost of data augmentation, data augmentation becomes a more common data augmentation method. In this study, data is extended by a mosaic data enhancement method. Mosaic data enhancement refers to the CutMix method of image stitching, which uses four images at a

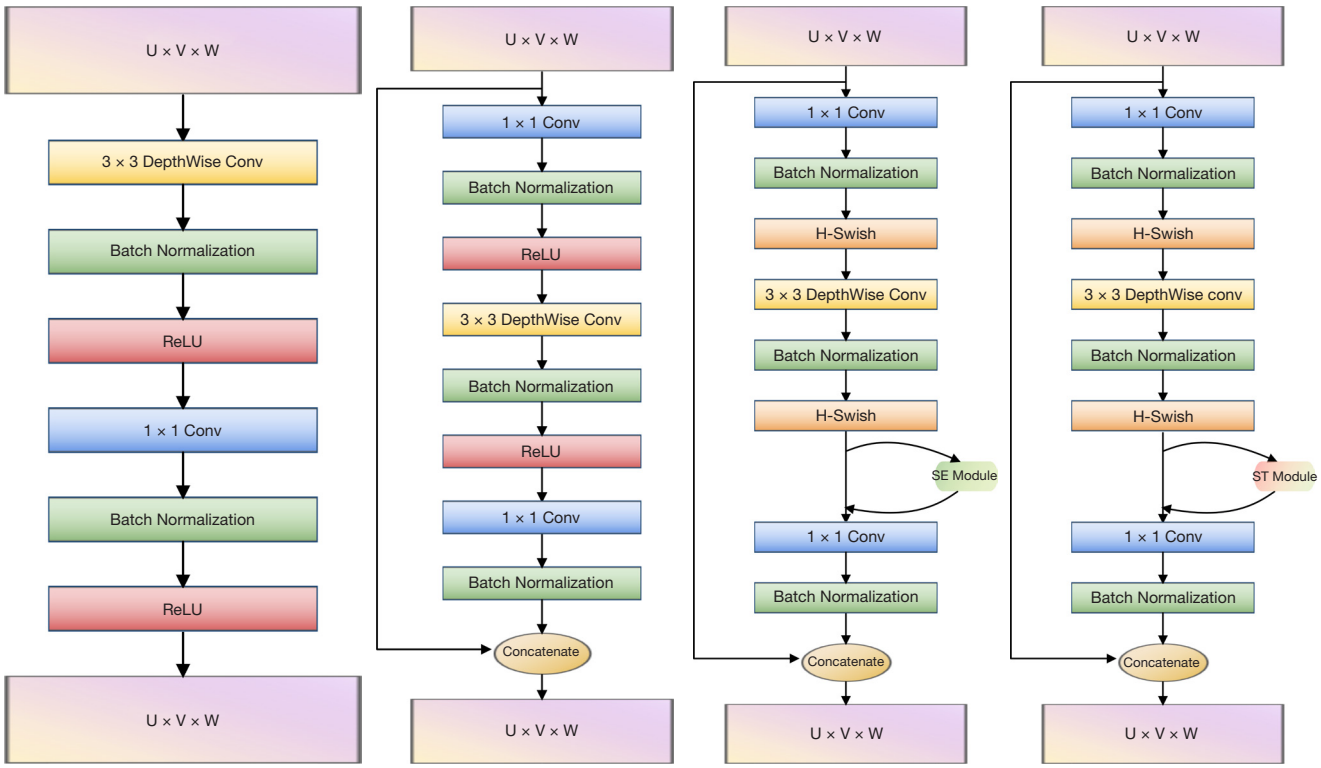

Figure 10 MobileNetV1, MobileNetV2, MobileNetV3, MobileNetV3+. Conv, convolution; SE, squeeze and excitation; ST, spatial-temporal attention mechanism.

time, flipping, scaling and changing the color gamut for each image, arranging them according to their positions in four directions. Then, it performs the combination of images and predicted frames.

Model training

To find a better performance model for lung nodule identification and detection, this study also proposed four different improvements to the backbone network of YOLOv4 using the MobileNet series (19) of lightweight networks. MobileNet family network structure and the improved mobilenetv3+ structure are shown in *Figure 10*. Specifically, mobilenetv1, mobilenetv2, and mobilenetv3 are used to replace the backbone network CSPDarknet53 of YOLOv4, and the replaced backbone networks are named YOLOv4-Mv1, YOLOv4-Mv2, and YOLOv4-Mv3, respectively. To reflect the space-time attention mechanism, the channel attention mechanism in mobilenetv3 was replaced with the spatial-temporal attention mechanism proposed in this paper, and then the backbone network of YOLOv4 was replaced, and the improved network was named YOLOv4-Mv3+. The backbone network constructed by G-BNeck was improved and finally named YOLOv4-GNet. YOLOv4 was used as the benchmark experiment for model selection. The model structure with the best performance was selected by experimental comparison, and the loss function was improved by using CIOU loss and cross-entropy loss function, respectively, to obtain better detection performance. After selecting the best network, we calculate the performance of the dataset on the best network and use our dataset to conduct experiments on some of the current mainstream target detection networks to compare the experimental results with our network.

The designed network model was trained based on a dataset from a tertiary hospital in Yunnan, China. The network model is implemented on PyTorch (version 1.5.0), and the experimental configuration for training and prediction is NVIDIA GeForce RTX2080Ti graphics card with 11G of video memory, 62G of RAM, and 100G of hard disk. the confidence level is set to 0.5, the Iou threshold is set to 0.5, and the decay weight coefficient is set to 0.0005. Ninety percent of the data are divided into training set and the other ten percent as the test set, where the training set

Table 1 Parameter counts and evaluation metrics of models

Model	Parameter count	Maximum GPU memory	mAP0.5	Average number of false positives	FPS
YOLOv4	40967325	244	34.50	56	52
YOLOv4-Mv1	12692029	69.7	38.91	47	67
YOLOv4-Mv2	10062013	53.6	51.38	32	72
YOLOv4-Mv3	13989484	91	29.01	68	89
YOLOv4-Mv3+	11729069	87.8	45.10	43	69
YOLOv4-GNet	11428545	64.9	54.56	27	70

GPU, graphics processing unit; mAP, mean average precision; FPS, frames per second.

continues to be divided randomly, 90% for training and 10% for validation. Meanwhile, in order to accelerate the learning speed of the network, a part of the network was first frozen for training, with the learning rate set to 0.001 and the batch size set to 4. Then the network was unfrozen and the training was continued, at which time the learning rate was set to 0.00001 and the batch size was set to 2, in order to achieve optimization of the model.

Diagnostic performance of the machine learning models

Evaluation metrics are used as a benchmark for comparing algorithms on the same data set, and to judge the performance of an algorithm or parameter settings. In this research, the model performance is evaluated and analyzed in terms of the number of model parameters, precision, sensitivity(recall), specificity, F1-score, maximum GPU memory, mAP, free-response receiver operating characteristic curve (FROC), challenge performance metric (CPM), and frames per second (FPS). Among them, parameter count can be loosely defined as the total parameter volume, except the number of fully connected layers. In deep learning, the parameter count represents the model size and the number of unit connections (computational cost) between layers.

The lower the parameter count, the lower the computational cost and the less memory the model needs. FROC indicates the relationship between the average number of false alarms detected per sample and sensitivity, and the horizontal coordinate indicates the average number of false positives and the vertical coordinate is the sensitivity, which can visually reflect the performance of the model. CPM is the average recall under the average number of false positives, and the higher the score, the better the performance of the model. The details of each evaluation criterion are as follows.

$$Precision = \frac{TP}{TP + FP} \quad [7]$$

$$\text{Sensitivity}=Recall = \frac{TP}{TP + FN} \quad [8]$$

$$Specificity = \frac{TN}{TN + FP} \quad [9]$$

The F1-score is a combined accuracy and recall measure equal to the harmonic mean of accuracy and recall. A higher F1-score indicates more efficient detection performance. Specifically, the F1-score can be calculated as follows:

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall} \quad [10]$$

The mAP refers to the average of the AP (average precision) values of all classes, and the AP value refers to the average precision of the predictions of a class.

First, we conduct experiments with our own dataset and take some evaluation metrics to demonstrate the superiority of the algorithms in this study. According to *Table 1*, the final algorithm used in this study is much lower than the original YOLOv4 network in terms of the number of parameters and the maximum GPU memory requirement, and lower than other improved networks except YOLOv4-Mv2. Although the number of parameters and memory requirements are increased compared to the YOLOv4-Mv2 network, the algorithm we finally used performs better in terms of detection accuracy and average false positive nodes. Among the proposed improved algorithms, our final adopted algorithm has the highest average recognition accuracy and the lowest average number of false positive nodes (the average number of false positive nodes is the ratio of the total number of false positive nodes to the total

number of CT images). Also, the efficiency of the algorithm can be compared more intuitively by using frames per second as the unit of consideration for operating speed. This shows that the modified YOLOv4-GNet model has different degrees of speed improvement compared to the initial YOLOv4 network and other modified networks. In the comparison between YOLOv4-MV3 and YOLOv4-MV3+, it was found that the spatial-temporal attention mechanism module reduced the computational speed, but the final algorithm YOLOv4-GNet algorithm still showed a significant improvement in computational speed compared with the original YOLOv4 algorithm. In contrast, the average number of false-positive nodes more than doubled.

Discussion

Although existing network models have shown promising object recognition performance, they either lack generality in their training scheme or sacrifice computational efficiency. Our model is different from those approaches in two aspects. First, our lightweight neural network considers both recognition accuracy and efficiency. The use of depthwise separable convolution significantly reduces the number of model parameters and calculation consumption, allowing us to achieve satisfying results in real-time with fewer training parameters. The compact structure also allows our model to be converged in a short time during training. Second, we propose a lightweight CNN model based on attention mechanism for image detection. The utilization of channel-wise and spatial-temporal attention mechanism can boost the representational power of network, and maintain a large range of dependencies, which allows our network to find cues near the lung nodule region. Therefore, our model has the potential for applications in real scenarios.

In order to illustrate the advantages of the proposed lightweight neural network for lung nodule detection, we apply our own dataset, perform experiments on landmark algorithms (R-CNN, SPP-Net, Fast R-CNN, Faster R-CNN, FPN, etc.) for object detection, and compare the results. R-CNN was proposed by Girshick *et al.* (20) It combines the region proposal and CNN, called region with CNN features, and its workflow is to generate class-independent candidate regions by selective search and then convolutional search. The workflow is to generate class-independent region proposals by selective search, then extract features by CNN, and finally output the results by linear SVM classifier. However, R-CNN requires a fixed-size image input, which will reduce the recognition accuracy of the original image or subgraph. Thus He *et al.* (21) proposed SPP-Net, which adds a "spatial pyramid pooling" pooling strategy to the R-CNN to eliminate the problem of fixed-size input. The network can eliminate the limitation of fixed-size input. At the same time, the region proposals can be extracted at the level of feature map to avoid the computational redundancy of candidate region feature extraction. However, both R-CNN and SPP-Net, the region proposals of their detection methods only provide rough locations. So Girshick (22) proposed Fast R-CNN based on R-CNN, which performs single-stage training and uses multi-task loss to update all network layers to improve detection accuracy and speed. Although Fast R-CNN improves the speed of object detection, the extraction of candidate regions is still obtained in advance with external algorithms, which cannot meet the real-time requirements and is not fully end-to-end training in the sense. Therefore, Ren *et al.* (23) proposed Faster R-CNN, which introduces region proposal network (RPN), which shares the convolutional features of the whole graph with the detection network, making the computation of region proposals take almost no extra time. However, considering that the computational energy consumption caused by the traditional feature pyramid computation is too large, Lin *et al.* (24) proposed to construct a feature pyramid using the multi-scale pyramid inherent in deep convolutional networks, named FPN, which can construct high-level semantic feature mappings at all levels of scales and can be trained end-to-end, and the network is used consistently during training and testing.

It is clear that the YOLOv4 network is the best performing network among the existing studies. In this research, the network is improved and its performance metrics are calculated. In order to test the performance of our network in image recognition, we selected some CNN models with good performance, and conducted some comparison experiments based on our own dataset. As shown in *Table 2*, the F1, precision, sensitivity, and specificity of the Yolov4-GNet network proposed in this research are 0.87, 86.34%, 86.69% and 90.63%, respectively. On our dataset, the detection performance of the Yolov4-GNet network proposed in this research is better than the current neural networks.

It is necessary to compare with the algorithms of other researchers in order to reflect the advances of the algorithm proposed in this paper. To facilitate the comparison, we

Table 2 Comparison of detection performance of YOLOv4-GNet and other models

Model	F1-score	Precision (%)	Sensitivity (%)	Specificity (%)
R-CNN	0.39	56.25	30.25	33.63
SPP-Net	0.26	68.95	16.03	20.75
Fast R-CNN	0.45	58.31	36.83	39.98
Faster R-CNN	0.56	70.00	47.06	46.88
FPN	0.57	83.00	43.70	46.87
Yolov4-GNet	0.87	86.34	86.69	90.63

R-CNN, region-convolutional neural network; SPP, spatial pyramid pooling; CNN, convolutional neural network; FPN, feature pyramid networks.

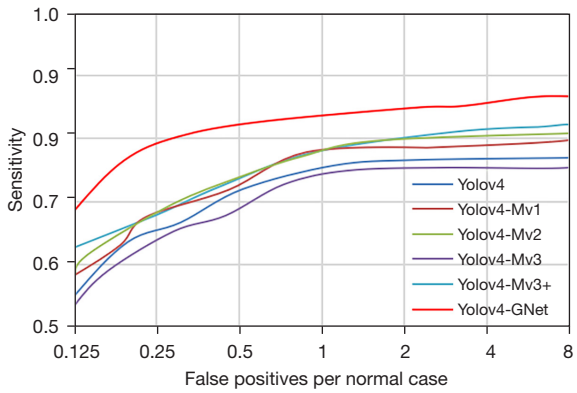

Figure 11 FROC value for different algorithms. FROC, free-response receiver operating characteristic curve.

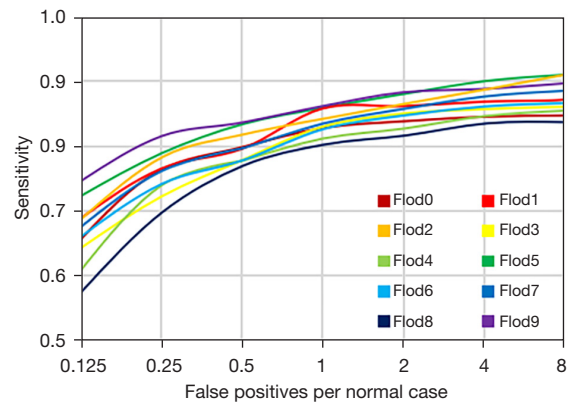

Figure 12 Ten-fold cross validation FROC chart. FROC, free-response receiver operating characteristic curve.

chose the public dataset LUNA16 for our experiments. The LUNA16 dataset is a subset of the LIDC dataset collected by the National Cancer Institute (NCI), which is one of the recognized sources of experimental data for studying lung lesion detection. In this section, the FROC and the CPM are used to compare the superior performance of the algorithms. The FROC is the relationship between the average number of false positives detected per sample and the sensitivity, where the horizontal coordinate is the average number of false positives and the vertical coordinate is the sensitivity, which can visually reflect the performance of the model. To ensure comparability, all algorithms in the following figure are studies with LUNA16 as the experimental dataset.

The FROC values of the different improved methods are shown in *Figure 11*, where the red line indicates the FROC curve of YOLOv4-GNet, which has higher sensitivity than other improved algorithms at each false positive rate point. Meanwhile, a ten-fold cross-validation of the algorithm was performed as shown in *Figure 12*. The results show that our algorithm has stable sensitivity scores at all false-positive rate points. It further illustrates the correctness of the improvement direction of the algorithm used in this research and the superiority of this algorithm.

In addition, we compared the detection sensitivity of different false-positive rates of pulmonary nodules with the existing papers, as shown in *Table 3*. The improved model outperforms some more recent papers and demonstrates good performance. Xie *et al.* (25) designed a 2D CNN-based boosting architecture for automatic lung nodule detection based on Faster R-CNN networks. The method was able to identify latent pulmonary nodules, but the method did not share the computational process in the final classification, leading to a considerable computational effort. Zuo *et al.* (26) proposed a multi-resolution CNN to classify pulmonary nodules. The network was able to identify less obvious nodules caused by radiological heterogeneity, but its extraction of contextual information was inadequate, resulting in an

Table 3 Comparison of detection sensitivity of different lung nodule detection algorithms

Algorithm	0.125	0.25	0.5	1	2	4	8	CPM
Xie [2019] (25)	73.40	74.40	76.30	79.60	82.40	83.20	83.40	79.00
Zuo [2019] (26)	67.20	69.40	71.40	73.90	76.60	78.70	82.20	74.20
You [2019] (27)	62.40	64.30	67.70	72.80	80.00	85.80	90.00	74.70
Sun [2021] (28)	46.7	60.2	73.0	81.2	87.7	92.0	93.1	76.20
Hong [2021] (29)	60.05	70.18	80.01	84.39	89.99*	92.16*	94.42*	81.60
Yang [2021] (30)	56.40	67.40	75.70	78.20	83.50	87.60	91.20	77.20
Improved model	69.83*	80.00*	82.36*	84.42*	84.79	85.23	86.16	81.83*

*, the highest sensitivity of the detection at that FP rate. CPM, challenge performance metric; FP, false positive.

overall sensitivity maintained in a low-level region. You *et al.* (27) designed a 3D CNN for screening pulmonary nodules, which achieved pulmonary nodule classification by reusing features using single-connected paths through constant mapping and residual unit acceleration model training. However, this network is limited to accepting only single-size sample input. Sun *et al.* (28) proposed a 3D attention embedded complementary flow CNN for lung nodule feature extraction and classification, which extracts background information about nodules using a progressive multi-scale feature extraction block with an attention module. However, this network does not take into account the data imbalance problem and performs poorly in terms of low false positive rate. Hong *et al.* (29) proposed a 3D CNN for lung nodule detection based on U-NET, which improves the quality of feature mapping generated by the lung nodule network through spatial and channel attention mechanisms. The network achieved good results at high error rates, but poor sensitivity at low false-positive rates, resulting in low overall CPM. Yang *et al.* (30) proposed a dense neural network-based nodule false-positive screening model, which enhanced feature utilization and expanded feature space through dense connections. However, this method failed to utilize the spatial features of nodules, so the CPM values achieved were not excellent. In contrast, although our algorithm has average sensitivity performance at an average of 2–8 false positives per scan, it has higher sensitivity than other current advanced algorithms at an average of 0.125–1 false positives per scan, which indicates that the present system can detect more nodules at a small error rate, i.e., it has high sensitivity at a lower scan FP rate. This can provide more direct help to physicians with the guaranteed low error rate, and has positive implications for improving the automation system of existing computer-aided diagnosis systems. The CPM in this research outperforms other algorithms, further proving the effectiveness of the algorithm.

Conclusions

In this research, we improve a deep learning-based object detection network for lung nodule images. This network not only shortens the examination time, but also reduces the examination burden on physicians, which its cost-effectiveness may help facilitate the use of computer-aided design in imaging examinations. Overall, the network constructed in this research has two main advantages. On the one hand, compared with the traditional dichotomy between malignant and benign, the method improves the sensitivity and accuracy of intelligent detection of pulmonary nodules at a low false positive rate, which is of great importance for computer-aided diagnosis. On the other hand, deep learning-based object detection algorithms have good generalization ability, and a spatial-temporal attention mechanism is employed to improve the performance of intelligent detection, so it is easier to detect some atypical nodules.

Although the development of computer technology has brought new techniques and results in the detection and diagnosis of lung nodules, there are still many limitations and clinical difficulties. The application of efficient, accurate and valuable computer-aided design models relies on a large number of samples. The sample size included in this research is small. Further expansion of the sample size and optimization of the model architecture, analysis of images, and textural features of typical benign nodules are still necessary to further optimize and validate the diagnostic model. In the next step, we will continue to

deepen our cooperation with hospitals to apply the model to clinical lung nodule detection experiments in order to bring help to clinical diagnosis. Due to the absolute privacy of medical image data, the owner of the data should take high protection measures to guarantee the information security when exchanging big data. Federated learning is a good solution to the privacy security problem and carries out efficient image recognition among multiple participants or multiple computational nodes. In future work, we will combine multi-angle images to diagnose lung nodule images and incorporate more advanced machine learning methods (e.g., federal learning) to enable better integration of research with clinical diagnosis.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (No. 71764035 and No. 71864021) and Yunnan Fundamental Research Projects (No. 202101AU070167).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-21-1182/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of the Yunnan Cancer Hospital. The requirement for written informed consent was waived due to the retrospective nature of the study.

References

1. Kikano GE, Fabien A, Schilz R. Evaluation of the Solitary Pulmonary Nodule. Am Fam Physician 2015;92:1084-91.
2. Brahmer JR, Govindan R, Anders RA, Antonia SJ, Sagorsky S, Davies MJ, et al. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of non-small cell lung cancer (NSCLC). J Immunother Cancer 2018;6:75.
3. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, Mehta AC, Ohno Y, Powell CA, Prokop M, Rubin GD, Schaefer-Prokop CM, Travis WD, Van Schil PE, Bankier AA. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. Radiology 2017;284:228-43.
4. Chu C, Zheng J, Zhou Y. Ultrasonic thyroid nodule detection method based on U-Net network. Comput Methods Programs Biomed 2021;199:105906.
5. Niu S, Huang J, Li J, Liu X, Wang D, Wang Y, Shen H, Qi M, Xiao Y, Guan M, Li D, Liu F, Wang X, Xiong Y, Gao S, Wang X, Yu P, Zhu J. Differential diagnosis between small breast phyllodes tumors and fibroadenomas using artificial intelligence and ultrasound data. Quant Imaging Med Surg 2021;11:2052-61.
6. Shen X, Wang L, Zhao Y, Liu R, Qian W, Ma H. Dilated transformer: residual axial attention for breast ultrasound image segmentation. Quant Imaging Med Surg 2022;12:4512-28.
7. Ma J, Song Y, Tian X, Hua Y, Zhang R, Wu J. Survey on deep learning for pulmonary medical imaging. Front Med 2020;14:450-69.
8. Li WJ, Lv FJ, Tan YW, Fu BJ, Chu ZG. Benign and malignant pulmonary part-solid nodules: differentiation via thin-section computed tomography. Quant Imaging Med Surg 2022;12:699-710.
9. Lee S, Kouzani AZ, Hu EJ. Automated detection of lung nodules in computed tomography images: a review. Machine Vision and Applications 2012;23:151-63.
10. Choi WJ, Choi TS. Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. Comput Methods Programs Biomed 2014;113:37-54.
11. Setio AA, Jacobs C, Gelderblom J, van Ginneken B. Automatic detection of large pulmonary solid nodules in thoracic CT images. Med Phys 2015;42:5642-53.
12. Javaheri T, Homayounfar M, Amoozgar Z, Reiazi R, Homayounieh F, Abbas E, et al. CovidCTNet: an open-source deep learning approach to diagnose covid-19 using

small cohort of CT images. NPJ Digit Med 2021;4:29.
13. Hu L, Zhou DW, Guo XY, Xu WH, Wei LM, Zhao JG. Adversarial training for prostate cancer classification using magnetic resonance imaging. Quant Imaging Med Surg 2022;12:3276-87.
14. Yang D, Ren G, Ni R, Huang YH, Lam NFD, Sun H, Wan SBN, Wong MFE, Chan KK, Tsang HCH, Xu L, Wu TC, Kong FS, Wáng YXJ, Qin J, Chan LWC, Ying M, Cai J. Deep learning attention-guided radiomics for COVID-19 chest radiograph classification. Quant Imaging Med Surg 2023;13:572-84.
15. Thakur SK, Singh DP, Choudhary J. Lung cancer identification: a review on detection and classification. Cancer Metastasis Rev 2020;39:989-98.
16. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016:779-788.
17. Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PY, Mu TJ, Zhang SH, Martin RR, Cheng MM, Hu SM. Attention mechanisms in computer vision: A survey. Computational Visual Media 2022;8:331-68.
18. Han K, Wang Y, Tian Q, Guo J, Xu C, Xu C. Ghostnet: More features from cheap operations. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020:1580-9.
19. Howard A, Sandler M, Chen B, Wang W, Chen LC, Tan M, Chu G, Vasudevan V, Zhu Y, Pang R, Adam H, Le Q. Searching for mobilenetv3. Proceedings of the IEEE/CVF International Conference on Computer Vision 2019:1314-1324.
20. Girshick R, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. IEEE Computer Society, 2014.
21. He K, Zhang X, Ren S, Sun J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans Pattern Anal Mach Intell 2015;37:1904-16.
22. Girshick R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision 2015:1440-1448.
23. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137-49.
24. Lin TY, Dollar P, Girshick R, He K, Hariharan B, Belongie S, editors. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017, pp. 936-44.
25. Xie HT, Yang DB, Sun NN, Chen ZN, Zhang YD. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. Pattern Recognition 2019;85:109-19.
26. Zuo WX, Zhou FQ, Li ZX, Wang L. Multi-Resolution CNN and Knowledge Transfer for Candidate Classification in Lung Nodule Detection. IEEE Access 2019;7:32510-21.
27. You K, Hao P, Wu F, Zhang F, Wu J. False positive screening of pulmonary nodules based on three-dimensional convolutional neural network. Journal of Graphics 2019;40:423-8.
28. Sun L, Wang Z, Pu H, Yuan G, Guo L, Pu T, Peng Z. Attention-embedded complementary-stream CNN for false positive reduction in pulmonary nodule detection. Comput Biol Med 2021;133:104357.
29. Hong M, Wu G, Liu X, Jia J, Yang X. Lung nodule detection algorithm based on attention mechanism. Computer Engineering and Design 2021;42:83-8.
30. Yang J, Zhang C, Dai J, Hao J, Wang S. False positive screening model of pulmonary nodules based on dense neural network. Computer Technology and Development 2021; 31:147-52.

Cite this article as: Yang L, Cai H, Luo X, Wu J, Tang R, Chen Y, Li W. A lightweight neural network for lung nodule detection based on improved ghost module. Quant Imaging Med Surg 2023;13(7):4205-4221. doi: 10.21037/qims-21-1182

