Lung Nodule Detection Using Multi-Scale Convolutional Neural Network and Global Channel Spatial Attention Mechanisms

Pham Minh Hieu
University of Science and Technology of Hanoi

***Abstract*—Early detection of lung nodules is a critical factor in lung cancer treatment. However, current methods still face many challenges such as missing small nodules, diversity in nodule sizes, and high false positive rates. In this study, I re-implement the lung nodule detection method based on Global Channel Spatial Attention Mechanism (GCSAM) proposed by Li et al. (2025). The system consists of two stages: Candidate Nodule Detection Network (CNDNet) and False Positive Reduction Network (FPRNet). CNDNet uses Res2Net as the backbone for multi-scale feature extraction, combined with GCSAM to adaptively adjust feature weights. FPRNet significantly reduces the false positive rate by accurately classifying true nodules and similar structures. Due to computational resource constraints, I trained the model on a subset of the LUNA16 dataset with 3 subsets for training and 1 subset for validation over 20 epochs using an A100 GPU on Google Colab. Experimental results demonstrate the method's capability for nodule detection with acceptable accuracy.**

***Index Terms*—lung nodule detection, 3D CNN, attention mechanism, deep learning, LUNA16**

I. Introduction

A. Background

Lung cancer is one of the most common and deadliest cancers worldwide, with approximately 1.3 million deaths annually [1]. According to the American Cancer Society statistics for 2023, an estimated 238,340 new cases of lung cancer were diagnosed in the United States, with 127,070 deaths from this disease [2].

The average 5-year survival rate for lung cancer patients is only about 16%, but early diagnosis and treatment can increase this rate to 70% [3]. In the early stages, lung cancer primarily manifests as lung nodules, and early detection of malignant nodules is the most effective method to prevent lung cancer.

B. Challenges

Computed Tomography (CT) has become an effective tool for detecting lung nodules, providing clear and intuitive images of lung lesions. However, detecting lung nodules on CT images still faces many challenges:

- **Large data volume:** Each CT scan generates hundreds of slices, significantly increasing the workload for radiologists and potentially leading to missed small nodules.
- **Morphological diversity:** Nodules have great diversity in size, shape, and density, often attached to blood vessels or bronchial walls, easily leading to misdiagnosis.
- **Small nodules:** Nodules with small sizes (3-5mm) are very difficult to detect due to low contrast with surrounding tissue.
- **High false positive rate:** Many normal structures (blood vessels, scar tissue) have characteristics similar to nodules, causing many false positives.

C. Research Objectives

This research aims to:

1) Re-implement a two-stage lung nodule detection system based on GCSAM.
2) Evaluate the model's performance on a subset of the LUNA16 dataset with limited computational resources.
3) Analyze the performance of each stage (CNDNet and FPRNet) in the detection pipeline.

II. Related Work

A. Traditional Methods

Traditional Computer-Aided Detection (CAD) systems primarily rely on hand-crafted feature extractors to identify candidate nodules [4], [5]. However, these systems typically only extract low-level features such as shape and texture, unable to effectively handle the diversity in size, shape, and density of lung nodules.

B. Deep Learning Methods

With the development of deep learning, methods based on Convolutional Neural Networks (CNN) have been widely applied in medical imaging:

- **2D CNN:** Fu et al. [6] combined handcrafted and deep features for nodule detection. Xie et al. [7] used Faster R-CNN with two region proposal networks.
- **3D CNN:** Ding et al. [8] proposed a Faster R-CNN network based on deconvolution architecture. Cao et al. [9] developed a Two-Stage CNN (TSCNN) model using U-Net and ensemble learning.
- **Multi-scale methods:** Tang et al. [10] proposed multi-scale feature 3D U-Net. Zheng et al. [11] developed a multi-scale feature detection network.
- **Attention mechanisms:** Li et al. [12] integrated SE-Net into residual blocks. Zhao et al. [13] proposed a 3D CNN model combining channel and spatial attention with Feature Pyramid Network.

C. Implemented Method

I re-implement the method proposed by Li et al. (2025) [14], including:

- Global Channel Spatial Attention Mechanism (GCSAM) for adaptive feature weight adjustment.
- Res2Net as backbone for multi-scale feature extraction.
- Hierarchical Progressive Feature Fusion (HPFF) to combine deep semantic information with shallow positional information.
- Two-stage system: CNDNet for candidate detection and FPRNet for false positive reduction.

III. Dataset and Data Split

A. LUNA16 Dataset

This study uses the LUNA16 dataset [**?**], which is constructed from the LIDC-IDRI database [**?**]. The dataset contains:

- 888 CT scans
- 1,186 annotated nodules with diameter $\geq$ 3mm
- 754,975 candidate nodules provided in *candidates.csv*, among which only 1,557 are true nodules, resulting in severe class imbalance

The dataset is officially divided into 10 subsets for cross-validation.

B. Data Split Strategy

Due to computational resource constraints (Google Colab with NVIDIA A100 GPU but I have limited compute unit :)), only a portion of the dataset was used in this study:

- **Training:** Subsets 1, 3, and 5
- **Validation:** Subset 7
- **Test:** Subset 9 and not performed on the full 10-fold cross-validation setup

This configuration corresponds to approximately 30% of the full dataset and does not follow the official 10-fold evaluation protocol of LUNA16. Therefore, the reported results should be interpreted as preliminary experimental findings rather than benchmark-level performance.

IV. Methodology

A. System Overview

The proposed lung nodule detection system consists of two main stages:

1) **Stage 1 - Candidate Nodule Detection:** Uses CNDNet to detect all regions likely to be nodules (high sensitivity).
2) **Stage 2 - False Positive Reduction:** Uses FPRNet to classify candidates into true nodules and false positives (high precision).

B. Stage 1: Candidate Nodule Detection Network

1) Network Architecture: CNDNet is designed based on an encoder-decoder architecture with main components:

a) Res2GCSA Module: This is the basic building block of the network, combining Res2Net [15] with GCSAM. This module performs:

- **Multi-scale feature extraction:** Splits feature maps into 4 subsets and processes them through different 3×3×3 convolution paths to extract features at multiple scales.
- **Channel attention:** Uses global context to calculate attention weights for each channel.
- **Spatial attention:** Adjusts features along spatial dimensions based on max pooling and average pooling.

b) Hierarchical Progressive Feature Fusion: HPFF is designed to combine features from different scales:

$$F_{fused} = \text{Concat}(F_4, F_3, F_2) \tag{1}$$

where F_i is the feature map from stage i after upsampling to size 32×32×32.

c) Region Proposal Network: RPN uses 3 anchor sizes (5mm, 10mm, 20mm) to detect nodules at different sizes. Each anchor box predicts:

$$\text{Output} = [p, \Delta x, \Delta y, \Delta z, \Delta d] \tag{2}$$

where p is the confidence score and $(\Delta x, \Delta y, \Delta z, \Delta d)$ are regression terms.

2) Loss Function: The loss function for Stage 1 includes classification loss and regression loss:

$$L = L_{cls} + \lambda L_{reg} \tag{3}$$

Classification Loss: Uses Binary Cross-Entropy:

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{4}$$

Regression Loss: Uses Smooth L1 Loss:

$$L_{reg} = \frac{1}{N_{pos}} \sum_{i \in \text{positive}} \text{SmoothL1}(t_i - t_i^*) \tag{5}$$

where:

$$\text{SmoothL1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \tag{6}$$

C. Stage 2: False Positive Reduction Network

1) Network Architecture: FPRNet is a 3D binary classifier with a simpler architecture than CNDNet:

- **Input:** 32×32×32 patches cropped from candidate nodule coordinates.
- **Encoder:** 3 stages of Res2GCSA blocks with MaxPooling and Dropout.
- **Fully Connected Layers:** 2 FC layers with Dropout for classification.
- **Output:** 2 class probabilities (nodule vs non-nodule).

2) Handling Class Imbalance: The candidate dataset has severe class imbalance between positive and negative samples (ratio 1:500). To address this, I apply:

- **Positive sample augmentation:** In the original study, positive samples were augmented 20 times. However, due to training time constraints, in my implementation, no augmentation was applied (factor = 0).
- **Negative sample downsampling:** Only keep 3% of negative samples.
- **Weighted loss:** Use class weights in the loss function.

3) Loss Function: Uses Weighted Cross-Entropy Loss:

$$L_{FPR} = -\frac{1}{N}\sum_{i=1}^{N}[w_1 y_i \log(p_i) + w_0(1-y_i)\log(1-p_i)] \quad (7)$$

where $w_1 = 10.0$ is the weight for positive class and $w_0 = 1.0$ for negative class.

D. Preprocessing

CT scan data is preprocessed through the following steps:

1) **Lung mask extraction:**
   - Threshold at -600 HU
   - Connected component analysis to remove non-lung regions
   - Erosion and hole filling
   - Convex hull processing
   - Binary dilation (10 iterations)
2) **HU clipping and normalization:**

$$I_{norm} = \frac{\text{clip}(I, -1200, 600) - (-1200)}{600 - (-1200)} \times 255 \quad (8)$$

V. Experimental Setup

A. Training Configuration

1) Training Environment:

- **Platform:** Google Colab
- **GPU:** NVIDIA A100 (40GB VRAM)
- **Framework:** PyTorch 1.8.1
- **Python:** 3.8.8

TABLE I
HYPERPARAMETERS FOR CNDNET

Parameter	Value
Input size	128×128×128
Batch size	4
Initial learning rate	0.01
Learning rate schedule	Step decay
Momentum	0.9
Weight decay	1×10^{-4}
Optimizer	SGD
Epochs	20
Anchor sizes	[5, 10, 20] mm

2) Hyperparameters - Stage 1:

3) Hyperparameters - Stage 2:

TABLE II
HYPERPARAMETERS FOR FPRNET

Parameter	Value
Input size	32×32×32
Batch size	8
Initial learning rate	0.01
Learning rate schedule	Step decay
Momentum	0.9
Weight decay	1×10^{-4}
Optimizer	SGD
Epochs	20
Negative sample ratio	3%
Positive class weight	10.0
Dropout rate	0.3

B. Evaluation Metrics

- **Free-Response Receiver Operating Characteristic (FROC):** Evaluates sensitivity according to false positives per scan.
- **Competition Performance Metric (CPM):** Average sensitivity at various FPs/scan:

$$CPM = \frac{1}{7}\sum_{i\in\{0.125,0.25,0.5,1,2,4,8\}} \text{Sensitivity}_{fps=i} \quad (9)$$

- **Sensitivity:**

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (10)$$

VI. Results and Discussion

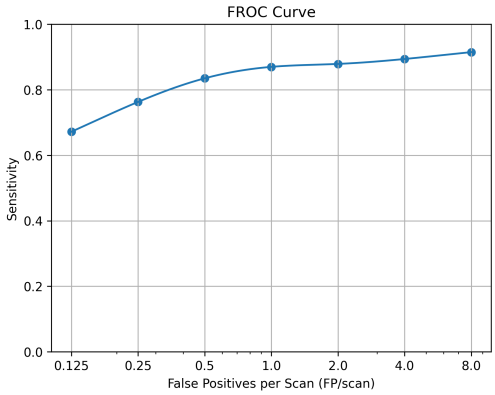

Fig. 1. FROC Curve

Figure 1 presents the FROC curve. The sensitivity increases steadily as FP/scan increases, with strong improvement in the low-FP region and gradual saturation afterward. The overall CPM score of 0.833 indicates stable detection performance under different false positive constraints.

Compared to many other studies, my result is quite humble. This is because several implementation modifications were applied: only 3/10 subsets were used, training was limited to 20 epochs, preprocessing excluded lung mask extraction, and no positive augmentation was

TABLE III
SENSITIVITY AT DIFFERENT FPS/SCAN - STAGE 1

FPs/scan	Sensitivity
0.125	0.672
0.25	0.763
0.5	0.835
1	0.870
2	0.879
4	0.894
8	0.915
CPM	**0.833**

applied in Stage 2. These changes reduce computational cost but may affect generalization and convergence.
Limited data and fewer training epochs may prevent full model convergence and reduce robustness to rare nodule patterns. The absence of lung masking may increase false positives from non-lung regions, while the lack of positive augmentation exacerbates class imbalance.
Despite these limitations, the two-stage architecture with attention mechanisms demonstrates reasonable detection capability. However, further training with the full dataset, extended epochs, and complete preprocessing is necessary before clinical deployment.

VII. CONCLUSION

A. Summary

In this study, I re-implemented a two-stage lung nodule detection system based on Global Channel Spatial Attention Mechanism. Due to computational resource constraints, I trained on a subset of the LUNA16 dataset (3 subsets for training, 1 subset for validation) over 20 epochs using an A100 GPU on Google Colab.
The system consists of two stages:

- **Stage 1 (CNDNet):** Detects candidate nodules with high sensitivity
- **Stage 2 (FPRNet):** Reduces false positives through binary classification

B. Contributions

Main contributions of this research:

a) Successfully implemented GCSAM and Res2Net architecture for lung nodule detection.
b) Evaluated model performance with limited training data and computational resources.
c) Provided baseline implementation that can be extended for future research.
d) Detailed analysis of the impact of data and training time limitations.

C. Limitations

This study has the following limitations:

- **Training data:** Only used 30% of dataset (3/10 subsets)
- **Training time:** 20 epochs instead of 200 epochs (Stage 1) and 50 epochs (Stage 2)
- **Preprocessing:** Did not perform full lung mask extraction
- **Data augmentation:** No positive sample augmentation in Stage 2
- **Validation:** Not yet evaluated on full test set

D. Future Work

To improve system performance and practicality, the following research directions can be pursued:

a) **Full dataset training:**
- Train on all 10 subsets of LUNA16
- Increase epochs to optimal level (200 for Stage 1, 50 for Stage 2)
- Apply 10-fold cross-validation as in original study

b) **Improve preprocessing:**
- Implement full lung mask extraction pipeline
- Experiment with different segmentation methods
- Optimize HU window settings

c) **Handle class imbalance:**
- Apply positive sample augmentation (factor = 20)
- Try Focal Loss instead of Weighted Cross-Entropy
- Optimize class weights based on validation performance

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA: A Cancer Journal for Clinicians*, vol. 65, no. 1, 2015.
[2] R. L. Siegel, K. D. Miller, N. S. Wagle, et al., "Cancer statistics, 2023," *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, 2023.
[3] N. Camarlinghi, "Automatic detection of lung nodules in computed tomography images: training and validation of algorithms using public research databases," *European Physical Journal Plus*, vol. 128, p. 110, 2013.
[4] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Medical Image Analysis*, vol. 14, no. 3, pp. 390–406, 2010.
[5] L. Lu, et al., "Hybrid detection of lung nodules on CT scan images," *Medical Physics*, vol. 42, no. 9, pp. 5042–5054, 2015.
[6] L. Fu, et al., "Automatic detection of lung nodules: false positive reduction using convolution neural networks and handcrafted features," in *Proc. SPIE Medical Imaging*, vol. 10134, pp. 60–67, 2017.
[7] H. Xie, et al., "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, pp. 109–119, 2019.
[8] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Proc. MICCAI 2017*, pp. 559–567, 2017.
[9] H. Cao, et al., "A two-stage convolutional neural networks for lung nodule detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 2006–2015, 2020.

[10] S. Tang, M. Yang, and J. Bai, "Detection of pulmonary nodules based on a multiscale feature 3D U-Net convolutional neural network of transfer learning," *PLoS ONE*, vol. 15, no. 7, p. e0235672, 2020.

[11] S. Zheng, et al., "A lower false positive pulmonary nodule detection approach for early lung cancer screening," *Diagnostics*, vol. 12, no. 11, p. 2660, 2022.

[12] Y. Li and Y. Fan, "DeepSEED: 3D squeeze-and-excitation encoder-decoder convolutional neural networks for pulmonary nodule detection," in *Proc. IEEE ISBI 2020*, pp. 1866–1869, 2020.

[13] Y. Zhao, et al., "Pulmonary nodule detection based on multiscale feature fusion," *Computational and Mathematical Methods in Medicine*, vol. 2022, p. 8903037, 2022.

[14] Y. Li, L. Hui, X. Wang, L. Zou, and S. Chua, "Lung nodule detection using a multi-scale convolutional neural network and global channel spatial attention mechanisms," *Scientific Reports*, vol. 15, no. 1, p. 12313, 2025.

[15] S. H. Gao, et al., "Res2Net: A new multi-scale backbone architecture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 652–662, 2019.