

Detecting lung nodules in computed tomography images based on deep learning

Lam Thanh Hien¹, Le Anh Tu², Pham Trung Hieu³, Pham Minh Duc⁴,
Nguyen Van Nang³, Do Nang Toan³

¹Faculty of Information Technology, Lac Hong University, Bien Hoa, Vietnam

²Faculty of Information Technology, Ha Long University, Uong Bi, Vietnam

³Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

⁴Computer Vision and Robotics Laboratory, International School, Vietnam National University, Hanoi, Vietnam

Article Info

Article history:

Received Mar 5, 2025

Revised Jul 19, 2025

Accepted Sep 14, 2025

Keywords:

CT image

Deep learning

Fine-tuned faster R-CNN

Lung nodule detection

Lung segmentation

ABSTRACT

Lung cancer is currently recognized as one of the most dangerous cancers, with high mortality rate. In order to deal with lung cancer, an important task is to detect lung nodules early to improve patient survival rates, and computed tomography (CT) scans are crucial data for this. In this research, we propose a deep learning-based method for detecting lung nodules in the CT images with the goal of increasing the likelihood of nodule appearance in the input data of the network, making it easier for the model to focus on relevant areas while reducing noise from areas unrelated to the result. Specifically, we propose a simple lung region segmentation process and optimize the hyperparameters of the faster region-based convolutional neural networks (faster R-CNN) model based on the analysis of nodule characteristics in CT image data. In our experiments, to evaluate the effectiveness of our proposals, we conducted tests on the standard LUNA16 dataset with different backbone configurations for the model, namely ResNet50, ResNet50v2, and MobileNet. The best results achieved were 0.86 mAP50 and 0.91 Recall for the ResNet50, and 0.84 mAP50 and 0.94 Recall for the ResNet50v2. These impressive outcomes underscore the success of our method and establish a robust basis for future studies to further integrate AI into healthcare solutions.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Do Nang Toan

Institute of Information Technology, Vietnam Academy of Science and Technology

18 Hoang Quoc Viet, Cau Giay District, Hanoi10072, Vietnam

Email: dntoan@ioit.ac.vn

1. INTRODUCTION

One of the leading causes of death worldwide for both women and men is lung cancer [1]. This is a serious illness characterized by a high rate of mortality, mainly because it is often diagnosed at a late stage when metastasis has occurred or there is severe damage to other organs in the body. However, early detection of lung cancer can lead to a substantial increase in the patient's survival rate [2]. One of the earliest signs of lung cancer is the appearance of round or oval-shaped tumors or nodules in the lungs [3]. These tumors generally vary in size, ranging from just a few millimeters to several centimeters in diameter. The tumor or lung nodule may have irregular or unclear borders and sometimes exhibit a surrounding shadow. The incidence of indeterminate lung nodules has been steadily increasing over the past few years [4]. According to the University of Texas Southwestern Medical Center [5], about 40% of lung nodules are cancerous. As

mentioned in the data from the World Health Organization (WHO) [6], approximately 1.8 million deaths occurred due to lung cancer in 2020.

A lung nodule is defined as a shadowed area, which can be round or irregular in shape, with a diameter ranging from 3 mm to 3 cm, and located within the air-filled lung [7]. Shadows with a diameter smaller than 3 mm are referred to as micronodules, while shadows with a diameter larger than 3 cm are classified as tumors [8]. It is worth noting that according to [5], about half of patients treated for lung nodules with cancer can survive at least five years after diagnosis. Furthermore, if the lung nodule is 1 cm or less in diameter, then the patient's five-year survival rate can rise to as high as 80%. That is why the task of early detection of lung nodules is so important.

An important tool for the task of detecting lung nodules is computed tomography (CT) imaging. CT images have the ability to provide detailed visuals of the internal structures of the body, which helps doctors identify abnormalities, including lung nodules. CT images offer high resolution and multiple slices, making it possible to determine the location, size, and shape of the lung nodules. Therefore, CT images are also crucial data for information technology research aimed at the automatic detection of lung nodules, thereby helping to accelerate and improve the accuracy of medical diagnoses.

The study in this paper focuses on the issue of detecting lung nodules in CT images using a deep learning approach. In deep learning, convolutional layers within convolutional neural networks (CNNs) are particularly effective at capturing spatial features from images, including CT scans. This is a key advantage for identifying lung nodules in CT scans, where challenges arise in distinguishing regions due to complex structural conditions and contrast variations between regions. Additionally, convolutional neural networks have been proven to be effective tools in various machine vision tasks, with specific network configurations and application conditions depending on the characteristics of the data.

In this study, the focus is on developing a process for detecting lung nodules in CT images with the goal of enhancing the likelihood of nodule appearance in the network's input data, allowing the model to more easily focus on relevant areas while reducing noise from regions unrelated to the result. Specifically, the study involves fine-tuning the faster region-based CNN (Faster R-CNN) model within the context of lung nodule data conditions in CT images. We also address the development of image preprocessing steps to improve the quality of the detection results. The main contributions of the paper include: i) Enhancing the effectiveness of lung nodule detection by proposing a lung region segmentation process to help deep learning models focus on processing the lung area; ii) Optimizing the hyperparameters of the Faster R-CNN model based on the analysis of lung nodule characteristics in CT image data; and iii) Clarifying the effectiveness of the proposals by implementing and evaluating the model with different backbones, specifically ResNet50, ResNet50v2, and MobileNet. The structure of the remaining paper is as: section 2 covers related works, section 3 presents the data used and the proposed methodology of the paper, Section 4 discusses the experiments and the evaluation of the results obtained. Finally, the conclusion is presented.

2. RELATED WORKS

One study on lung nodule detection in CT images is mentioned in [9], where the research introduces a method that uses filters designed to enhance the quality of dot-like and line-like objects in two-dimensional space. The study uses sensitivity and specificity metrics to evaluate the performance of the filters. results on the low dose CT (LDCT) dataset show that 93.4% of 76 nodules were detected, with a false positive rate of 4.2. The study [10] discusses an automated method for detecting lung nodules through multiple image preprocessing steps and the use of wavelet transform and biorthogonal wavelet techniques to enhance image quality. First, the bi-histogram equalization algorithm is used to balance the contrast of the image. Next, morphological transformations are applied to clean and segment the image, helping to separate important structures in the lungs. The segmented regions are then extracted and fed into a fuzzy inference system (FIS). This FIS is used to determine the severity of lung nodules, aiding in the classification and evaluation of tumors. This method was tested on data from 25 patients. The results show that the method has the ability to detect lung cancer early, confirming the potential of wavelet transformation techniques, Bi-Histogram Equalization, and morphological filters in the automatic identification of lung nodules.

In addition to the studies mentioned above, there are several studies that use deep learning to detect lung nodules in CT images. In 2016, a study [11] was conducted to detect lung nodules in CT images using a 3D CNN. This method applies deep learning techniques to analyze complex medical image data, specifically chest CT images. The dataset used in the study is the LIDC-IDRI dataset, which includes 1,018 chest CT images. The performance metrics used to evaluate the method include sensitivity, false positive rate, and the free-response receiver operating characteristic (FROC) curve. The results indicated that the system reached a sensitivity of 78.9%, with an average of 20 false positives per scan (FPs/scan). This demonstrates that the system has a good sensitivity for detecting lung nodules, although the number of false positives may be higher compared to some other methods. A noteworthy point is that this system does not use additional

segmentation processes or false positive reduction techniques, which may contribute to the higher number of false positives. However, the study confirmed the potential of the 3D CNN in lung nodule detection, opening up a new direction for improving automated diagnostic systems. [12] discusses a study on the development and evaluation of a computer-aided detection (CAD) system based on 2D CNN and CT image segmentation techniques to detect lung nodules. The dataset used in the study is the LIDC-IDRI dataset, which includes 1,018 chest CT images. To assess the performance of the CAD system, the study uses metrics such as sensitivity, accuracy, and FPs/scan. A free-response operating characteristic (FROC) analysis was also conducted to measure sensitivity and the FPs/scan rate, enabling a comparison of performance between different CAD systems. The results of the study show that the proposed CAD system achieved a sensitivity of 92.8% with 8 false positives per scan. This is a very promising result when compared to other studies, demonstrating that the system can detect lung nodules with high accuracy and fewer false positives. The study also compares the performance of different CNN architectures, including AlexNet, GoogLeNet, and ResNet. In 2021, a study [13] conducted in Vietnam used various convolutional neural networks (CNNs) such as ATT, ASS, and AST to detect tumors in lung CT images. The study used the LIDC-IDRI dataset for evaluation. Precision, recall, and specificity were employed to evaluate the effectiveness of the CNNs. The results of the study showed impressive outcomes, with a precision of 95%, recall of 86.4%, and specificity of 98.9%. This outcome represents a major step forward in developing automated techniques for tumor detection in lung CT scans, contributing to enhanced accuracy in diagnosis and improved treatment planning. Also in Vietnam, a method using the LdcNet network [14] was implemented to classify lung CT images from the LIDC-IDRI dataset. To assess the model's effectiveness, evaluation metrics including accuracy, sensitivity, and specificity were utilized. The results of the study showed accuracy, specificity, and sensitivity values of 97.2%, 97.3%, and 96.0%, respectively. This demonstrates that the LdcNet network achieved very high performance in classifying lung images from the LIDC-IDRI dataset. These results could provide significant benefits in supporting the diagnosis and treatment of lung diseases.

3. MATERIALS AND METHODS

3.1. Dataset used

In this study, the dataset used is LUNA16 [15], which consists of 888 CT slices. The LUNA16 dataset was developed from the lung image database consortium and image database resource initiative (LIDC-IDRI) dataset, a large database containing lung CT images for analysis. LIDC-IDRI is used for lung cancer diagnosis research, and LUNA16 is an extension of this project, focusing on lung nodule detection. Each image in LUNA16 has pixel values ranging from -1000 to 3000, measured in Hounsfield units (HU), which represent the density of tissues in the body. The LUNA16 dataset is widely used in studies on detecting malignant lung nodules from CT scan data. These slices have a relatively high spatial resolution, with a size of 512×512 pixels. Additionally, this dataset includes labels indicating the location and size of the lung nodules, which were annotated by several medical experts, specifically radiologists. This enables deep learning models to be trained and evaluated on accurate medical diagnostic data. In detail, there are a total of 1,186 lung nodules in the 888 CT images. Figure 1 illustrates a CT slice.

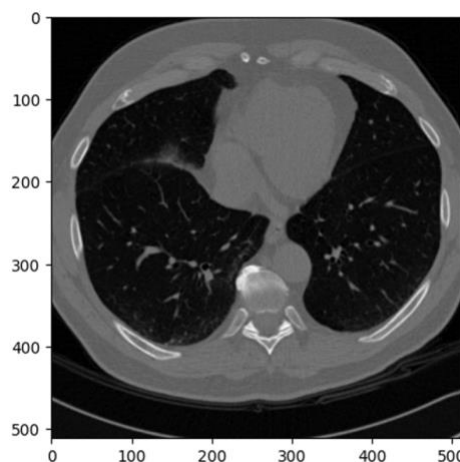


Figure 1. A CT slice of a lung from the LUNA16 dataset

3.2. Data preprocessing

3.2.1. Lung segmentation

According to the definitions of lung nodules presented in section 1, lung nodules are always fully located within the area of the two lungs. For this reason, accurately identifying the position of the two lungs in CT images becomes a crucial factor in detecting lung nodules. To enhance the ability to detect lung nodules, we propose a lung region segmentation technique, with the main goal being to separate the lung region from the surrounding areas. Lung segmentation allows the model to focus on important features and eliminate unnecessary areas, thereby improving the accuracy of locating lung nodules. The lung segmentation process is carried out in the following steps in Figure 2.

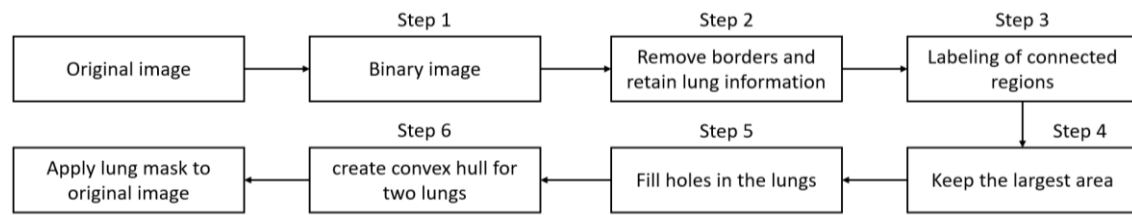


Figure 2. Diagram of the lung segmentation steps

Step 1 is converting the CT image into a binary image by using a specific threshold. In this case, the threshold value is selected as -604 HU. Pixels with values lower than this threshold are assigned a value of 1, and the remaining pixels are assigned a value of 0 (background). The pixels with a value of 1 are temporarily considered as the lung region. The algorithm takes a CT image and the threshold as input and returns the corresponding binary image. First, create an empty binary image of the same size as the input CT image. Then, iterate through each pixel, and if the pixel value is smaller than the threshold, assign a value of 1 to the corresponding pixel in the binary image; otherwise, assign a value of 0.

Next is step 2. After binarizing the image, the lung region is temporarily considered to be the area where the pixel values are equal to 1. However, areas outside the body will also have pixel values of 1, as this region is also air, like the lungs. Therefore, these areas need to be converted to background, corresponding to pixel values of 0. Step 2 is used to eliminate the boundary areas and retain the lung region. The algorithm takes a binary image and a padding size as input and returns the corresponding image with the boundary regions removed. First, we need to check if the padding size is negative; if it is, throw an exception. Then, add padding to the image with the corresponding size, and set the pixel values of the added padding to 1. Next, we use the algorithm by Wu *et al.* [16] to identify the connected components in the image obtained from the previous step. The algorithm by Kesheng Wu is fast and provides high accuracy for connected components. Then, the pixels at the boundary are assigned a value of 0 to become the background. The boundary area is defined as the area where at least one pixel touches the edge of the image (which could be at the top, bottom, leftmost, or rightmost edge). Finally, if padding was added in the previous step, the padding will be removed to return the image to its original size. The algorithm will return the corresponding image after the boundary regions have been removed.

Next is step 3. Label the connected components to identify the remaining foreground areas in the image after removing the boundary regions. Step 3 performs this task. First, Kesheng Wu's algorithm is used again to identify the connected components in the binary image. The result includes a list of regions, labels for each corresponding region, information about the size, and the coordinates of the bounding boxes of these regions. Based on the information from the bounding boxes, the connected components are drawn for a visual representation. The list of connected components is iterated through, and a corresponding bounding box is drawn around each region on the binary image. This is done by using a rectangle to surround each region, with the rectangle color set to green and the line width set to 2 pixels.

Next is step 4. It can be seen that removing the border might not necessarily keep only the lung information, as it could also contain other regions. For example, fat areas usually have low pixel values, so through the binary image process, these regions will have pixel values of 1, mixing with the areas containing air. The lung region is identified by calculating the area of each connected region and only keeping the two largest areas, corresponding to the two lungs.

Next is step 5. Filling the holes in the lung region helps ensure that the lung area is continuous and not fragmented, allowing the resulting lung mask to cover all the information within the lungs. Step 5 fills the holes within the lung area, which correspond to positions with pixel values of 0. First, the Canny algorithm [17] is applied to the mask to create a new image containing only the outlines. Then, morphological closing is

applied to close the small holes in the mask. The algorithm continues by finding connected regions in the mask and filling the small holes by comparing their area with the *max_hole_size* (a value that is predefined). The result is a lung mask with the small holes filled.

Next is step 6. The creation of a convex hull for the lungs helps eliminate small indentations around the surface of the lungs. First, we find the contours in the image using the algorithm by Satoshi Suzuki et al. [18]. Since non-lung areas or noise within the lungs have been removed in previous steps, only the two contours of the two lungs remain. Then, the algorithm iterates through each contour found in the previous step and computes the convex hull for each corresponding contour. This convex hull calculation is performed using Sklansky's algorithm [19]. Afterward, the pixels within the convex hull are assigned a value of 1 to create a lung mask.

Finally, the binary mask is overlaid on the original CT image to eliminate areas outside the lungs, retaining only the relevant lung regions. Through the lung segmentation steps above, the resulting image will contain only the two lungs, which allows models to focus exclusively on the lung area and enhances the ability to detect the location of lung nodules. Figure 3 illustrates each step of the proposed method.

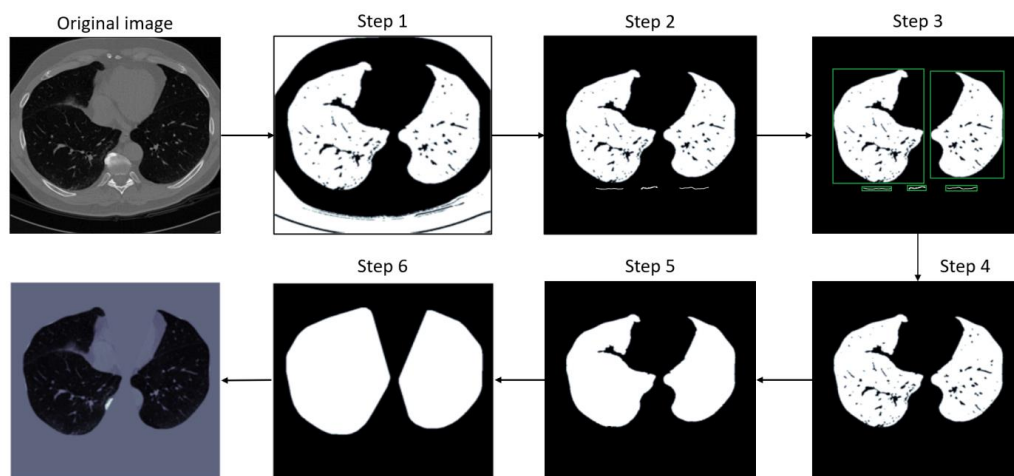


Figure 3. Illustrate the steps performed on a slice

3.2.2. Data normalization

Additionally, before the data can be input into the model, the CT image needs to be normalized. The pixel values may contain discrepancies due to errors from the imaging machine, and normalizing the image before feeding it into the model can help reduce these discrepancies and give the model an overall view of the dataset. The pixel values will be scaled to the range [0-1]. First, the minimum value (*min_val*) and maximum value (*max_val*) of the image are determined. Then, the pixel values will be normalized to the range [0-1] using the following formula:

$$norm_pixel = \frac{pixel - min_val}{max_val - min_val} \quad (1)$$

3.3. Optimizing the hyperparameters of the faster R-CNN

Faster R-CNN [20] has been extensively applied in medical image object recognition applications for several important reasons related to accuracy, the ability to detect complex objects, and the model's flexibility. In medical images, the objects that need to be detected are often very small or have complex shapes, such as lung nodules. Faster R-CNN, with its ability to learn and optimize features from multiple levels within the CNN network, can accurately detect and classify these objects. The Faster R-CNN architecture is composed of two primary components: the region proposal network (RPN) and the Fast R-CNN detection module.

The RPN plays a crucial role in the Faster R-CNN model, as it is responsible for generating candidate regions of interest from the input image, helping to identify areas that may contain objects. RPN operates based on a CNN to extract features from the image, then uses a sliding window to scan through points on the feature map. At each point, RPN generates "anchors" (assumed rectangles with different scales and aspect ratios) and classifies them into two groups: containing an object or not containing an object.

Simultaneously, the RPN predicts positional refinements for the anchors using a bounding box regression network, allowing them to more accurately align with the actual objects in the image. After generating the proposals, RPN uses the Non-Maximum Suppression (NMS) technique to eliminate overlapping regions and retain the best ones. Integrating RPN into deep learning models helps accelerate the object detection process, while simultaneously optimizing the generation of region proposals and object classification, thereby improving the efficiency and accuracy of the model.

In this study, we will optimize the parameters of the region proposal network module in the Faster R-CNN model based on the analysis of lung nodule characteristics in CT image data. The original architecture of Faster R-CNN predefines 3 different sizes [128, 256, 512] and 3 aspect ratios [0.5, 1, 2] for the anchor boxes. As a result, Faster R-CNN utilizes nine different anchor box sizes in total: [64, 128], [128, 128], [256, 128], [128, 256], [256, 256], [512, 256], [256, 512], [512, 512], and [1024, 512]. These predefined sizes are considerably larger than the typical diameter of lung nodules, resulting in a large number of unnecessary or irrelevant anchor boxes. Taking into account the actual sizes of lung nodules and the aspect ratio distribution in the LUNA16 dataset, we adopt three smaller anchor box sizes [32, 64, 128] combined with two aspect ratios [1, 2], resulting in six anchor box configurations: [32, 32], [32, 64], [64, 64], [64, 128], [128, 128], and [128, 256]. These anchor box sizes are smaller and better aligned with the actual dimensions of lung nodules in the dataset. Figure 4 presents a comparison between the original anchor boxes and the modified anchor boxes employed in our approach.

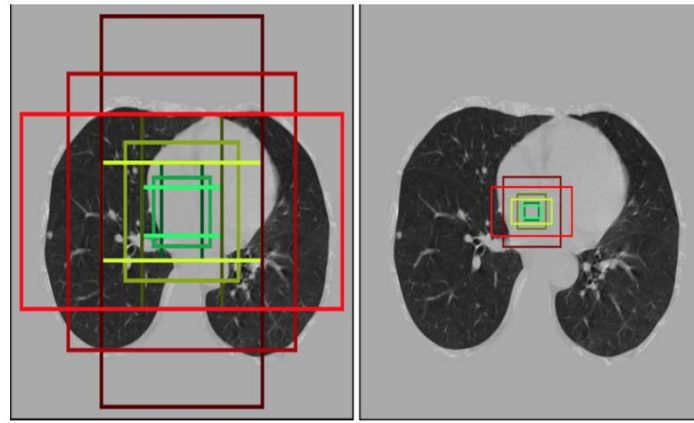


Figure 4. Comparison of Faster R-CNN using the original anchor boxes (left) and the customized anchor boxes tailored to lung nodules (right)

The Fast R-CNN detector in Faster R-CNN is the next step after the region proposals are generated by the RPN. Once the regions that may contain objects are identified, Fast R-CNN processes them to classify the objects and accurately determine their positions in the image. Fast R-CNN operates by using region proposals generated from the RPN and performs the following steps: first, these proposals are cropped from the feature map (the feature map extracted by the CNN from the input image). Then, a technique called region of interest (RoI) pooling is used to resize the proposals of different sizes into a fixed size, making them suitable for further processing in the network. These features are then passed into a fully connected (FC) network for object classification and bounding box coordinate prediction (position and size) for each proposal. The Fast R-CNN detector uses a multi-task loss function that combines classification loss and bounding box regression loss. The formula for the loss function is as:

$$L(p_i, t_i, v_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, v_i), \quad (2)$$

In the formula, N_{cls} denotes the number of RoIs used for classification, while N_{reg} refers to the number of RoIs used for bounding box regression. p_i is the predicted probability that the i^{th} RoI contains an object and p_i^* is the ground truth label, which takes values of 1 or 0 to indicate whether the i^{th} RoI is an object or not. t_i represents the ground truth bounding box parameters for the i^{th} RoI and v_i denotes the predicted bounding box parameters for the i^{th} RoI. L_{cls} is the classification loss function (typically cross entropy), and L_{reg} is the loss function for bounding box coordinate regression (usually smooth L1 loss). A balancing parameter λ is introduced to weight the contributions of the classification and regression losses in the overall loss function.

Fast R-CNN introduces several improvements over the previous method, R-CNN, including reducing computational time by processing the image only once through the CNN to extract features, rather than processing it multiple times for each proposal. This makes Fast R-CNN faster and more efficient in object detection. Additionally, Fast R-CNN can optimize the entire learning process, from feature extraction and classification to bounding box adjustment, resulting in a more powerful and accurate object detection model. In this study, we use three CNN backbone architectures in Faster R-CNN: ResNet50 [21], ResNet50 v2 [22], and MobileNet [23]. Figure 5 depicts the overall structure of the Faster R-CNN framework employed in this study.

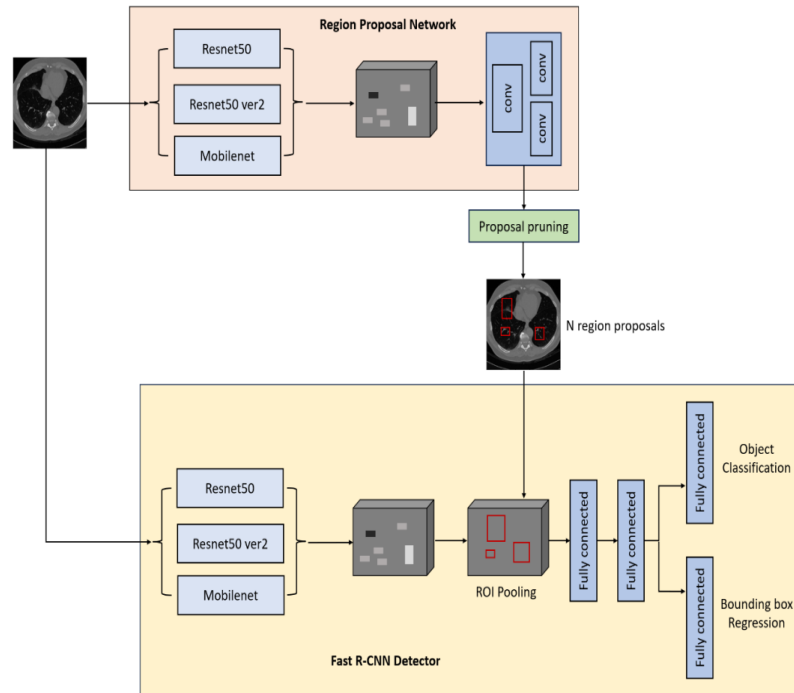


Figure 5. The architecture of Faster R-CNN with ResNet50, ResNet50 v2, and MobileNet backbones

4. EXPERIMENT AND RESULTS

4.1. Experiment

To clarify the effectiveness of the proposed method, we carry out experiments with two scenarios: using image data that includes both lungs and using image data that has been cropped to include only one lung. The use of one-lung image data is based on our hypothesis to enhance the performance of detecting lung nodules by directing the Faster R-CNN models to focus on processing the lung area. Therefore, the results with one-lung data are expected to perform better. To ensure a fair evaluation of the proposed method's effectiveness on the Faster R-CNN model, the metrics will be calculated based on the Faster R-CNN model with different backbone variations, including ResNet50, ResNet50 v2, and MobileNet.

In terms of experimental details, the model was trained using 2 T4 GPUs on the Kaggle platform [24]. The data was augmented using techniques such as image cropping, brightness adjustment, and blurring to increase the generalization of the dataset. The model was trained with a learning rate of 0.0001, a batch size of 10, and was run for 100 epochs. Additionally, from the preprocessed data, we performed image cropping to split each image into two new images corresponding to the two separate lungs. Thus, we will experiment and evaluate on two datasets. The first dataset is the uncropped dataset, which includes 1,186 CT image slices with a size of 512×512 . The second dataset is the one where each image is split into two halves according to the two lungs. To allow the model to focus more on images containing lung nodules, this dataset will consist of 100% images containing nodules and only 20% of images without nodules. Therefore, the dataset will consist of 1,304 images, each with a size of 256×384 . The test set of this dataset will be taken from the test set of the two-lung dataset after being split in half. The train, validation, and test sets are randomly split from the original dataset with a ratio of 7:2:1. Figure 6 illustrates an image with two lungs that has been cropped into two separate lung images.

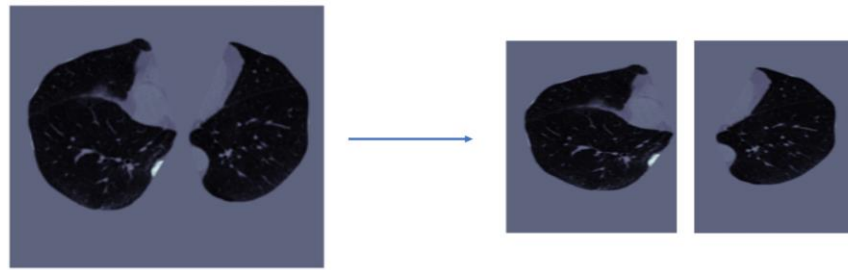


Figure 6. Illustration of a lung image split into two separate lung images

For this problem, the metrics used to evaluate the model are mAP50 (mean average precision at intersection over union (IoU) threshold of 50%) and recall. For a prediction result, evaluation is based on true positive (TP) and false positive (FP) through IoU. If the $\text{IoU} > 50\%$, the result is considered a TP, otherwise, it is not. If a correct result is not detected by the model, it is considered a false negative (FN).

In the deployment stage, we used the optimal model weights obtained from the experimental phase and built a graphical user interface to facilitate smooth interaction with the model. This desktop application was developed using Python's Tkinter library, which is well-suited for creating GUI applications. The system accepts a 2D CT image slice as input and processes it to detect and highlight any lung nodules present. The output, consisting of the original image with the identified nodule marked, is displayed within the interface using the Matplotlib library. For better visualization, the CT images are shown using a grayscale colormap, which is commonly applied in medical imaging to emphasize structural features.

4.2. Result and discussion

First, experiments were conducted using the dataset containing both lungs, the results are summarized in Table 1. The Faster R-CNN model with ResNet50, after being trained for approximately 100 epochs, achieved the best result with a recall of 0.8064 and mAP50 of 0.7. The loss function on the training set dropped below 0.1, indicating that the model achieved a relatively high sensitivity in detecting positive cases. However, the average precision (mAP50) of the predictions still needs improvement. Similarly, the Faster R-CNN model with ResNet50v2, after being trained for around 100 epochs, achieved the best result with a Recall of 0.83 and mAP50 of 0.74. This model produced the best results on the dataset with both lungs but, in return, the prediction time for an image was longer due to the architecture of the model. The Faster R-CNN model with MobileNet achieved the best result with a Recall of 0.75 and mAP50 of 0.63. This model converged quite quickly during training but had a lower Recall and mAP50 compared to the other two models. However, it performed predictions very quickly due to its smaller model structure.

Table 1. Experimental results of backbones of Faster R-CNN with the two-lung image dataset

Metrics	Resnet50	ResNet50v2	MobileNet
Recall	0.8064	0.83	0.75
mAP50	0.7	0.74	0.63

For experiments with the one-lung image dataset, the results show a clear difference as shown in Table 2. After training the model for approximately 100 epochs, the Faster R-CNN model with ResNet50 achieved a recall of 0.91 and mAP50 of 0.86. Splitting the two-lung image into individual lungs made nodule detection easier. The recall indicates that the model can accurately identify 91% of the nodules in the test dataset, and the mAP50 of 0.86 shows that the model quality is noticeably better compared to the two-lung image dataset. The Faster R-CNN model with ResNet50v2 achieved the best results with a recall of 0.94 and mAP50 of 0.84. While it can identify more accurately than the Faster R-CNN with the ResNet50 backbone, the lower mAP50 indicates that the model detects more false positives in the testset. The Faster R-CNN model with MobileNet, after training for around 100 epochs, achieved a Recall of 0.83 and mAP50 of 0.80. This model is the fastest in terms of prediction among all six models and performs better than the models using the two-lung image dataset.

To evaluate the results more objectively, we compared our results with the study in [25]. In that study, the authors presented a lung nodule detection pipeline using the YOLOv3 model with the Darknet53 backbone. The model was pre-trained on the MS COCO dataset. In experiments with the LUNA16 dataset,

the authors achieved their best results with a mAP50 of approximately 0.64 and a recall of around 0.75. The authors also compared the performance with several other models, including Cascade R-CNN with a mAP50 of 0.46 and a recall of 0.60, FCOS with a mAP50 of 0.50 and a recall of 0.72, and finally, YOLOv3 with the MobileNet backbone, which achieved a mAP50 of 0.51 and a recall of 0.83. Thus, compared to the best results we achieved, our models with ResNet50 and ResNet50v2 backbones demonstrated significantly better performance, with mAP50 values of 0.86 and 0.84, and Recall values of 0.91 and 0.94, respectively. These results clearly highlight the effectiveness of applying the proposed process for segmenting lung areas, as well as optimizing the region proposal network module in the Faster R-CNN model based on the analysis of lung nodule characteristics in CT image data. The proposed method enhances lung nodule detection by directing the model's attention toward relevant regions while effectively suppressing noise from irrelevant areas, thereby improving both accuracy and focus.

Table 2. Experimental results of backbones of Faster R-CNN with the one-lung image dataset

Metrics	Resnet50	ResNet50v2	MobileNet
Recall	0.91	0.94	0.83
mAP50	0.86	0.84	0.80

Our findings demonstrate that incorporating a lung region segmentation step prior to detection can effectively reduce background noise and enhance the model's focus on relevant anatomical areas. This preprocessing approach, although lightweight, contributes to improved performance in detecting lung nodules from CT scans. These results add to the current body of research by highlighting the combined importance of data preparation and model configuration. While previous studies have primarily focused on end-to-end detection frameworks, our work shows that even modest, domain-specific adjustments in preprocessing and model tuning can lead to substantial improvements.

Additionally, we have integrated all the components of this study into a testing software, as illustrated in Figure 7. The software allows the user to select one of the models studied in this paper and use chest CT images as input data. It then outputs the detected lung nodule regions along with their corresponding probabilities. This information is displayed visually on the program's interface.

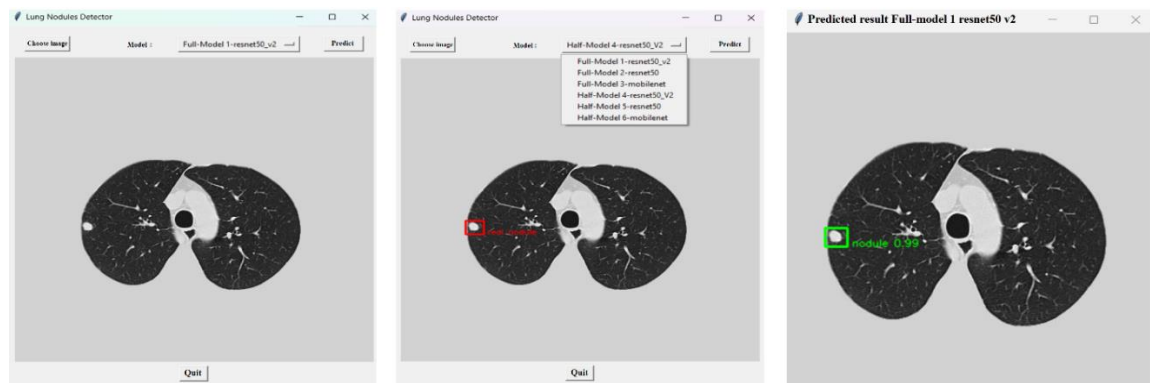


Figure 7. Illustration of the interface of the lung nodule detection software

5. CONCLUSION

In this paper, we presented a study on lung nodule detection in CT images using a deep learning approach, aiming to enhance the visibility of lung nodules and help the model focus more effectively on relevant regions while reducing noise from areas unrelated to the results. To demonstrate the effectiveness of the proposed algorithm, we conducted experiments on the standard LUNA16 dataset and achieved the best results of 0.86 mAP50 and 0.91 Recall for the ResNet50 backbone model, and 0.84 mAP50 and 0.94 Recall for the ResNet50v2 backbone model. These results reflect the effectiveness of the proposed method and highlight the potential application of the study in supporting lung cancer diagnosis. Despite achieving high evaluation metrics, the research still requires further improvement and testing in the context of real-world data, with more diverse and complex cases of lung CT images. Our future research will aim to improve the generalization capability of the employed deep learning architectures, while also investigating the 3D

structural characteristics of CT image data. Additionally, with the goal of translating the research into practical healthcare applications, we will work on specific tasks to integrate this method into the workflow of radiologists in hospitals, allowing for objective and practical assessments.

FUNDING INFORMATION

The research content in this paper is sponsored by the Institute of Information Technology (Viet Nam Academy of Science and Technology) under the project code “TXKHTE.02/25-25-02”.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Lam Thanh Hien	✓				✓	✓		✓		✓			✓	✓
Le Anh Tu				✓		✓		✓		✓	✓			
Pham Trung Hieu			✓				✓		✓		✓			
Pham Minh Duc			✓				✓							
Nguyen Van Nang				✓				✓						✓
Do Nang Toan	✓	✓			✓				✓			✓	✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in LUNA16 at <https://luna16.grand-challenge.org/Home/>

REFERENCES

- [1] K. V. Hinisha and A. Lijiya, “Lung nodule identification,” in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Jul. 2019, pp. 451–455, doi: 10.1109/ICICICT46008.2019.8993218.
- [2] D. E. Midthun, “Early detection of lung cancer,” *F1000Research*, vol. 5, p. 739, Apr. 2016, doi: 10.12688/f1000research.7313.1.
- [3] A. Panunzio and P. Sartori, “Lung cancer and radiological imaging,” *Current Radiopharmaceuticals*, vol. 13, no. 3, pp. 238–242, Nov. 2020, doi: 10.2174/1874471013666200523161849.
- [4] A. R. Larici *et al.*, “Lung nodules: size still matters,” *European Respiratory Review*, vol. 26, no. 146, p. 170025, Dec. 2017, doi: 10.1183/16000617.0025-2017.
- [5] “Long Nodule,” *UT Southwestern Medical Center*. <https://utswmed.org/conditions-treatments/lung-nodules> (accessed Dec. 12, 2024).
- [6] “Lung cancer,” *World Health Organization*. <https://www.who.int/news-room/fact-sheets/detail/lung-cancer> (accessed Dec. 12, 2024).
- [7] D. M. Hansell, A. A. Bankier, H. MacMahon, T. C. McLoud, N. L. Müller, and J. Remy, “Fleischner society: glossary of terms for thoracic imaging,” *Radiology*, vol. 246, no. 3, pp. 697–722, Mar. 2008, doi: 10.1148/radiol.2462070712.
- [8] M. Lederlin, M.-P. Revel, A. Khalil, G. Ferretti, B. Milleron, and F. Laurent, “Management strategy of pulmonary nodule in 2013,” *Diagnostic and Interventional Imaging*, vol. 94, no. 11, pp. 1081–1094, Nov. 2013, doi: 10.1016/j.diii.2013.05.007.
- [9] Q. Li, S. Sone, and K. Doi, “Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans,” *Medical Physics*, vol. 30, no. 8, pp. 2040–2051, Aug. 2003, doi: 10.1118/1.1581411.
- [10] C. C. Samuel, V. Saravanan, and M. R. V. Devi, “Lung nodule diagnosis from CT images using fuzzy logic,” in *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, Dec. 2007, pp. 159–163, doi: 10.1109/ICCIMA.2007.236.
- [11] R. Golan, C. Jacob, and J. Denzinger, “Lung nodule detection in CT images using deep convolutional neural networks,” in *2016 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2016, pp. 243–250, doi: 10.1109/IJCNN.2016.7727205.
- [12] Q. Wang, F. Shen, L. Shen, J. Huang, and W. Sheng, “Lung nodule detection in CT images using a raw patch-based convolutional neural network,” *Journal of Digital Imaging*, vol. 32, no. 6, pp. 971–979, Dec. 2019, doi: 10.1007/s10278-019-00221-3.
- [13] K. D. Lai, T. T. Nguyen, and T. H. Le, “Detection of lung nodules on CT images based on the convolutional neural network with attention mechanism,” *Annals of Emerging Technologies in Computing*, vol. 5, no. 2, pp. 78–89, Apr. 2021,

Detecting lung nodules in computed tomography images based on deep learning (Lam Thanh Hien)

- doi: 10.33166/AETiC.2021.02.007.
- [14] G. S. Tran, T. P. Nghiem, V. T. Nguyen, C. M. Luong, and J.-C. Burie, "Improving accuracy of lung nodule classification using deep learning with focal loss," *Journal of Healthcare Engineering*, vol. 2019, pp. 1–9, Feb. 2019, doi: 10.1155/2019/5156416.
 - [15] A. A. A. Setio *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, Dec. 2017, doi: 10.1016/j.media.2017.06.015.
 - [16] K. Wu, E. Otoo, and A. Shoshani, "Optimizing connected component labeling algorithms," in *Medical Imaging 2005: Image Processing*, Apr. 2005, vol. 5747, pp. 1965–1976, doi: 10.1117/12.596105.
 - [17] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.
 - [18] S. Suzuki and K. Be, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, Apr. 1985, doi: 10.1016/0734-189X(85)90016-7.
 - [19] J. Sklansky, "Finding the convex hull of a simple polygon," *Pattern Recognition Letters*, vol. 1, no. 2, pp. 79–83, Dec. 1982, doi: 10.1016/0167-8655(82)90016-2.
 - [20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
 - [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, doi: 10.1109/cvpr.2016.90.
 - [22] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9908, Springer International Publishing, 2016, pp. 630–645, doi: 10.1007/978-3-319-46493-0_38.
 - [23] A. G. Howard *et al.*, "MobileNets: efficient convolutional neural networks for mobile vision applications," *arXiv:1704.04861*, 2017.
 - [24] "Kaggle," *Kaggle*. <https://www.kaggle.com> (accessed Dec. 12, 2024).
 - [25] D. Fang, H. Jiang, W. Chen, Z. Qin, J. Shi, and J. Zhang, "Pulmonary nodule detection on lung parenchyma images using hyper-deep algorithm," *Heliyon*, vol. 9, no. 7, p. e17599, Jul. 2023, doi: 10.1016/j.heliyon.2023.e17599.

BIOGRAPHIES OF AUTHORS



Lam Thanh Hien    joined a M.Sc. in applied informatics at the INNOTECH Institute, France, and received a degree in 2004. In 2017, he earned a Ph.D. degree at the Vietnam Academy of Science and Technology. Now, he works at Lac Hong University in the headmaster's role. His studies interests relate to machine learning, computer vision, and deep learning. He can be contacted at email: is.lthien@lhu.edu.vn.



Le Anh Tu    graduated with a master's degree in 2007 from Thai Nguyen University, and a Ph.D. in 2017 from the Institute of Information Technology, Vietnam Academy of Science and Technology. He is currently a lecturer at Ha Long University. His main research areas are data mining, machine learning, and artificial neural networks. He can be contacted at email: leanhtu@daihochalong.edu.vn.



Pham Trung Hieu    learned computer and information science at Hanoi University of Science, Vietnam National University, and received a degree in 2023. He has continued to pursue a master's program in data science at Hanoi University of Science, Vietnam National University, from 2023 to the present. Now he is a researcher at the Vietnamese Academy of Science and Technology. His studies interests relate to data mining, deep learning, statistical model and applied mathematics in computer science. He can be contacted at email: pthieu@ioit.ac.vn.



Pham Minh Duc    received his bachelor's degree in computer science from the University of Science, Vietnam National University, Hanoi. He is currently working at the Computer Vision and Robotics Laboratory, International School, Vietnam National University. His research focuses on computer vision, including image processing, object recognition, motion tracking, and the application of deep learning models in image data analysis. He has been involved in various AI projects, particularly in developing and optimizing deep learning models for real-world computer vision applications. He can be contacted at email: pmduc2808@gmail.com.



Nguyen Van Nang    learned multimedia communication at the University of Information and Communications Technology - Thai Nguyen University, Thai Nguyen, and received a degree in 2017. Now, he is a researcher at the Vietnamese Academy of Science and Technology. His studies interests relate to machine learning, computer vision, and virtual reality technology. He can be contacted at email: nangnv@ioit.ac.vn.



Do Nang Toan    studied applied mathematics and informatics at Hanoi University and received a degree in 1990. In 2001, he earned a Ph.D. degree at the Vietnam Academy of Science and Technology. Now, he is an associate professor at the Vietnamese Academy of Science and Technology. His studies interests relate to machine learning, computer vision, and virtual reality. He can be contacted at email: dntoan@ioit.ac.vn.