

A DATA-CENTRIC DEEP LEARNING METHOD FOR PULMONARY NODULE DETECTION

CHI CUONG NGUYEN¹, LONG GIANG NGUYEN², GIANG SON TRAN^{1,*}

¹*ICTLab, University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Cau Giay District, Ha Noi, Viet Nam*

²*Institute of Information Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Cau Giay District, Ha Noi, Viet Nam*



Abstract. Lung cancer is one of the most serious cancer-related diseases in Vietnam and all over the world. Early detection of lung nodules can help to increase the survival rate of lung cancer patients. Computer-aided diagnosis (CAD) systems are proposed in the literature for early detection of lung nodules. However, most of the current CAD systems are based on the building of high-quality machine learning models for a fixed dataset rather than taking into account the dataset properties which are very important for the lung cancer diagnosis. In this paper, we follow the direction of data-centric approach for lung nodule detection by proposing a data-centric method to improve detection performance of lung nodules on CT scans. Our method takes into account the dataset-specific features (nodule sizes and aspect ratios) to train detection models as well as add more training data from local Vietnamese hospital. We experiment our method on the three widely used object detection networks (Faster R-CNN, YOLOv3, and RetinaNet). The experimental results show that our proposed method improves detection sensitivity of these object detection models up to 4.24%.

Keywords. Data-centric learning, deep learning, pulmonary nodule detection.

1. INTRODUCTION

Lung cancer is the leading deadliest cancer in both men and women worldwide [1]. In 2021, about 22% of cancer deaths in both genders in the US are lung cancers [1]. In Vietnam, lung cancer accounts for 182,563 new cases in 2020, equivalent to 14.4% of total new cancer-related cases [2]. Lung cancer is the second most popular cancer in Vietnam, only behind liver cancer. Statistics on survival rate of lung cancer vary depending on the stage of cancer when it is diagnosed. People have 60% of 5-year relative survival rates in localized stage (the cancer has not spread outside of the lung) and only 6% at distant stage (the cancer has spread to distant parts of the body) [3]. There is a greater chance that lung cancer can be cured effectively with early stage cancer.

In recent years, high-performance deep learning models are widely popular in computer-aided diagnosis (CAD) systems for early detection of pulmonary nodules in CT scan images [4–7]. In these approaches, a fixed lung nodule dataset is usually selected to fine-tune popular deep learning detection models such as reducing or adding some layers of neural

*Corresponding author.

E-mail addresses: cuongnc@st.usth.edu.vn (C.C. Nguyen); nlgiang@ioit.ac.vn (L.G. Nguyen); tran-giang.son@usth.edu.vn (G.S. Tran)

networks or using different loss functions when training the model, etc. This model-centric methods, however, may miss the important features of lung nodule datasets when training the detection model.

To fill the disadvantage of model-centric approach, one direction is to focus on data analysis and modification to support training a deep learning model. These research methods are called data-centric approach. The analysis and modification of data can be data augmentation, changes in ratio of training/validation/test sets, adding external resources, or improving dataset annotations, etc. In other words, while a model-centric method focuses on changing the model parameters and architectures to improve detection performance, a data-centric approach pays attention on improving the dataset quality for better overall performance.

In this work, we propose a new data-centric approach to improve performance of deep learning detection model on LUNA16 dataset [8], which is the most popular public dataset for lung nodule detection on CT scan images. In detail, our method takes into account the two important features of lung nodule datasets (nodule sizes and aspect ratios) to customize the architecture of deep learning detection model. We also collect more local Vietnamese lung nodule data from Vietnam National Cancer Hospital to enrich LUNA16 dataset.

Unlike most other data-centric deep learning methods mainly focus on data assimilation, data synthesizing, data preprocessing, and feature engineering [9], our contributions in this work are two-fold as follows:

1. Enriching LUNA16 dataset with adjacent slices and local Vietnamese data for lung nodule detection on CT scan images.
2. Improving performance of lung nodule detection models by taking into account two important features (nodule sizes and aspect ratios) of lung nodule dataset.

We evaluate the effectiveness of the proposed data-centric method on the three popular object detection methods (Faster R-CNN [10], YOLOv3 [11] and RetinaNet [12]). The experiments will show that our methods help in enhancing detection sensitivity of these object detection models up to 4.24%.

The rest of the paper is organized as follows. Section 2 introduces our method to enrich the LUNA16 dataset with additional data from local Vietnamese dataset (K dataset). Section 3 details our method to customize the detection models using lung nodule features. We present our experiments and discussions for the proposed methods in Section 4. Finally, Section 5 concludes this work and presents future research directions.

2. ANALYSIS OF LUNG NODULE FEATURES

2.1. LUNA16 dataset

The LUNA16 dataset was created for the Lung Nodule Analysis 2016 challenge [8]. It is currently the most widely used public dataset to solve the problem of early detection of lung nodules in CT scan images. LUNA16 dataset was extracted from the LIDC-IDRI dataset [13], which contains 1,018 thoracic CT scans gathered from several academic institutions. Each CT scan is connected with an XML file storing information of nodules and the corresponding annotations.

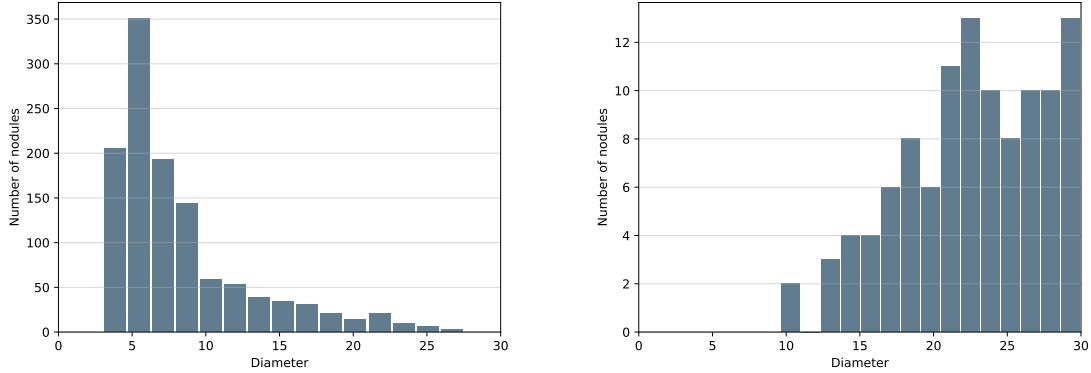


Figure 1: Distribution of nodule sizes in LUNA16 dataset (left) and K dataset (right)

The annotation in LIDC/IDRI dataset was generated by experienced radiologists to indicate if a lesion is nodule $\geq 3\text{mm}$, nodule $< 3\text{mm}$, or nonnodule (any other pulmonary abnormality). From this dataset, LUNA16 challenge extracted only nodules with the size between 3mm and 30mm to solve the problem of lung nodule detection on CT scan images. For those cases, the annotations contain position, diameter of nodule in each CT slice, and subjective ratings of the pathologic features such as calcification, internal structure, subtlety, lobulation, etc. [14].

To create a more consistent dataset, LUNA16 challenge discarded thin-slice CT scans with a slice thickness greater than 3mm and having inconsistent slice spacing or missing slice. After this operation, 888 CT scans are kept out of 1,018 cases of the LIDC/IDRI dataset. To obtain the lung nodules of high quality for the challenge, LUNA16 extracted only nodules marked by at least 3 out of 4 radiologists from LIDC/IDRI dataset. As a result, the challenge got 1,186 nodules as positive candidates for pulmonary nodule detection and classification.

2.2. K dataset

In the context of a scientific cooperation project between University of Science and Technology of Hanoi (USTH) and Vietnam National Cancer Hospital (K Hospital) from 2018 to 2019, we developed a Vietnamese lung nodule dataset, called K dataset. In this local Vietnamese dataset, we collected 382 CT scans from 382 Vietnamese patients from K hospital. These CT scans were gathered from the Siemens Emotion 6 scanner and Hitachi Scenaria 64-128 slices scanner in which all protected health information (PHI) from DICOM meta-data was removed to protect patient's information with Health Insurance Portability and Accountability Act (HIPAA) principles.

After having 382 CT scans from K hospital, we applied the same annotation protocol of LIDC/IDRI datasets to mark the lung nodules in the K dataset. Two experienced radiologists from Vietnam National Cancer Hospital were in charged of this annotation process in the context of the cooperation project. After the annotation process, we obtained 170 nodules of size between 3mm and 30mm out of 382 CT scans. We call this collection of 170 nodules as K dataset in this paper.

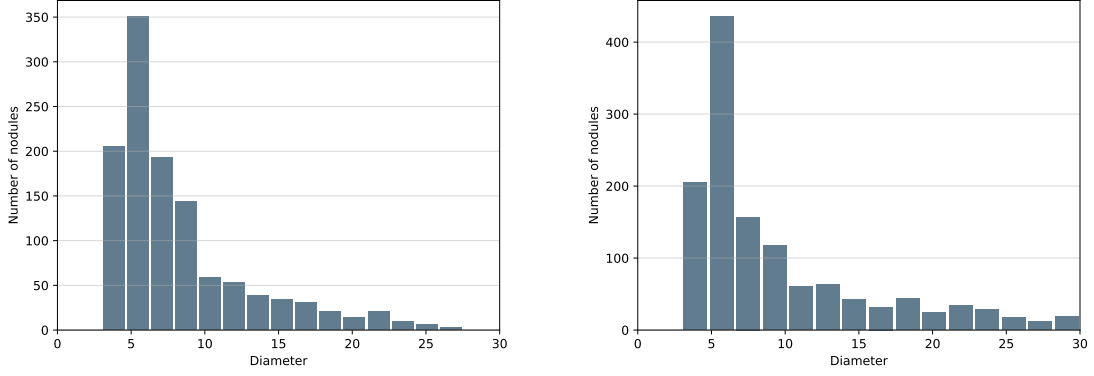


Figure 2: Distribution of nodule sizes in the original LUNA16 dataset (left) and in the enhanced LUNA16+K dataset (right). The new dataset is more balanced.

2.3. Analysis of lung nodule features

After constructing K dataset, we make an analysis of nodule sizes and aspect ratios in both K dataset and LUNA16 dataset for comparison. Fig. 1 shows this comparison. As can be seen from the figure, a majority of nodules in LUNA16 dataset has the size in the range from 3mm to 10mm (the histogram of the nodule sizes is deviated to the left). In contrast, most of nodules in K dataset has the size in the range from 15mm to 30mm (the histogram of the nodule sizes is deviated to the right).

From the above data analysis, we can infer that most current machine learning models detect the nodule sizes in the range from 3mm to 10mm with LUNA16 dataset while in our K dataset, most nodule sizes are in another range (from 15mm to 30mm). From this difference, we propose an enhanced version of LUNA16 dataset by integrating K dataset into LUNA16 dataset. By this integration, we can create a more nodule balanced dataset, called LUNA16+K (as seen in Figure 2). In the experiments, we will show that this LUNA16+K dataset helps in increasing the performance of lung nodule detection models.

3. INTEGRATING NODULE FEATURES TO IMPROVE PULMONARY NODULE DETECTION MODELS

To detect lung nodules on CT scans, most existing works [4, 15–21] employ the state-of-the-art object detection frameworks such as Faster R-CNN [10], YOLOv3 [11] and RetinaNet [12]. By default, these methods are designed to work with object detection on natural images, in which the objects are roughly centered and cover most of the space. To detect objects, these frameworks use a concept called anchors, which is a set of bounding boxes around the object. Each anchor contains two parameters: scale and aspect ratio, representing the relation of the width and the height of object candidates.

As presented in Section 2, pulmonary nodule candidates have the size in the range from 3mm to 30mm, which are quite small in comparison with objects in natural images. Besides, the number of anchor scales and their aspect ratios are important hyperparameters and highly depend on the specific features of each dataset. Taking this information into account,

we propose to adjust default anchor box sizes of the most popular object detection models (Faster R-CNN, YOLOv3 and RetinaNet) using the nodule features inside the dataset. More specifically, we customize default anchor boxes to get smaller ones, relatively close to nodule sizes as in Fig. 1 to train object detection models for lung nodule detection in CT scan images. Using smaller anchor box sizes should propose better and reliable bounding boxes and therefore, improve the detection performance. In the following subsections, we present in detail our customization of anchor box sizes of the experimented object detection frameworks: Faster R-CNN, YOLOv3, and RetinaNet.

3.1. Faster R-CNN

Faster R-CNN [10] is a two-stage detection method, which first proposes bounding box proposals then classifies each proposal as objects. The network is composed of three main components: A Feature Extraction Network (FEN), Region Proposal Network (RPN) and Region-of-Interest (ROI) classification. To train Faster R-CNN, input images are first fed into FEN, essentially a sequence of convolutional layers, for feature map extraction. These feature maps, along with pre-defined anchor boxes, are then put into RPN using a sliding window technique to generate region proposals, each is a bounding box indicating a potential region containing an object inside the input image. Each region proposal is then further classified as an object or background with the ROI classification network. Outputs of Faster R-CNN include object bounding boxes and the corresponding object class probability.

The original Faster R-CNN architecture predefines 3 different scales [128, 256, 512] and 3 symmetrical aspect ratios [0.5, 1, 2]. Therefore, Faster R-CNN has a total of 9 anchor box sizes: [64, 128], [128, 128], [256, 128], [128, 256], [256, 256], [512, 256], [256, 512], [512, 512] and [1024, 512]. These default sizes are significantly bigger than pulmonary nodule diameters, as can be illustrated on Fig. 3 (left), results in many irrelevant bounding boxes. Based on the real nodule size and ratio distribution on both LUNA16 and K dataset, we propose to use 3 smaller anchor boxes scales [32, 64, 128] with 2 aspect ratios [1, 2], resulting in a total of 6 anchor box sizes: [32, 32], [32, 64], [64, 64], [64, 128], [128, 128] and [128, 256]. The smaller anchor box sizes are more compatible with nodule sizes in the two datasets, as demonstrated on Fig. 3 (right).

3.2. YOLOv3

YOLOv3 (You Only Look Once) [11] was proposed as a single-stage object detection method. Unlike region-based methods, YOLO passes the image only once in a fully convolutional neural network (FCNN), making it relatively fast and realtime. It splits the image into grids, then predicts bounding boxes and their class probabilities for each region. YOLOv3 is an improvement and overcomes the performance of YOLOv1 [22] and YOLOv2 [23]. The first improvement of YOLOv3 is the use of bounding box and objectiveness score for each prediction. The second improvement is the use of Feature Pyramid Net (FPN) [24] to improve performance.

The most salient feature of YOLOv3 is the usage of 9 anchor boxes. YOLOv3 anchor configuration is applied as a result of K-means clustering method. For our datasets, this results in smaller anchor sizes [10, 13], [16, 30], [33, 23], [30, 61], [62, 45], [59, 119], [116, 90], [156, 198], and [373, 326]. Since there is no default anchor box configuration, we do not

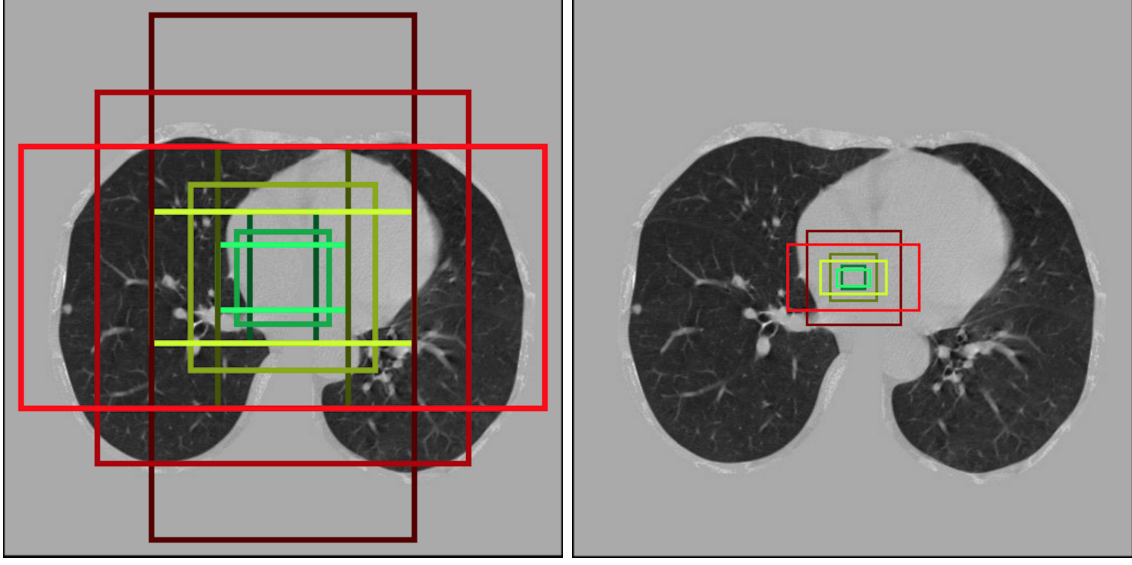


Figure 3: Faster R-CNN with original anchor boxes (left) and customized anchor boxes (right).

compare the benefit of customizing this setting with the original YOLOv3 implementation.

3.3. RetinaNet

RetinaNet [12] is one of the most state-of-the-art one-stage object detectors. It is composed of three main components: a backbone network for feature extraction, a feature pyramid network and a set of subnetworks for classification and bounding box regression. The backbone network is based on ResNet50 [25] to generate multi-scale feature maps from the input image. The feature pyramid network is used to perform upsampling of lower resolution feature maps (on higher pyramid levels) to higher ones (on lower pyramid levels) for better generalization of features. The two remaining subnetworks, box regression subnet and classification subnet, are responsible for predicting probability of an object in the corresponding anchor box and for regressing bounding box for each candidate.

By default, RetinaNet uses base anchor box sizes of [32, 64, 128, 256, 512] and aspect ratios [0.5, 1, 2]. For small pulmonary nodule detection, we halve these anchor sizes to [16, 32, 64, 128, 256] while keeping the same aspect ratios. These result in 15 anchor boxes of size [8, 16], [16, 16], [32, 16], [16, 32], [32, 32], [64, 32], [32, 64], [64, 64], [128, 64], [64, 128], [128, 128], [256, 128], [128, 256], [256, 256] and [512, 256].

4. EXPERIMENTS

4.1. Dataset preparation

We evaluate performance of our customized nodule detection models on two different datasets: LUNA16 (1,186 nodules) and a combination of LUNA16 with the K dataset (additional of 170 nodules). Each nodule candidate is provided with bounding box coordinates

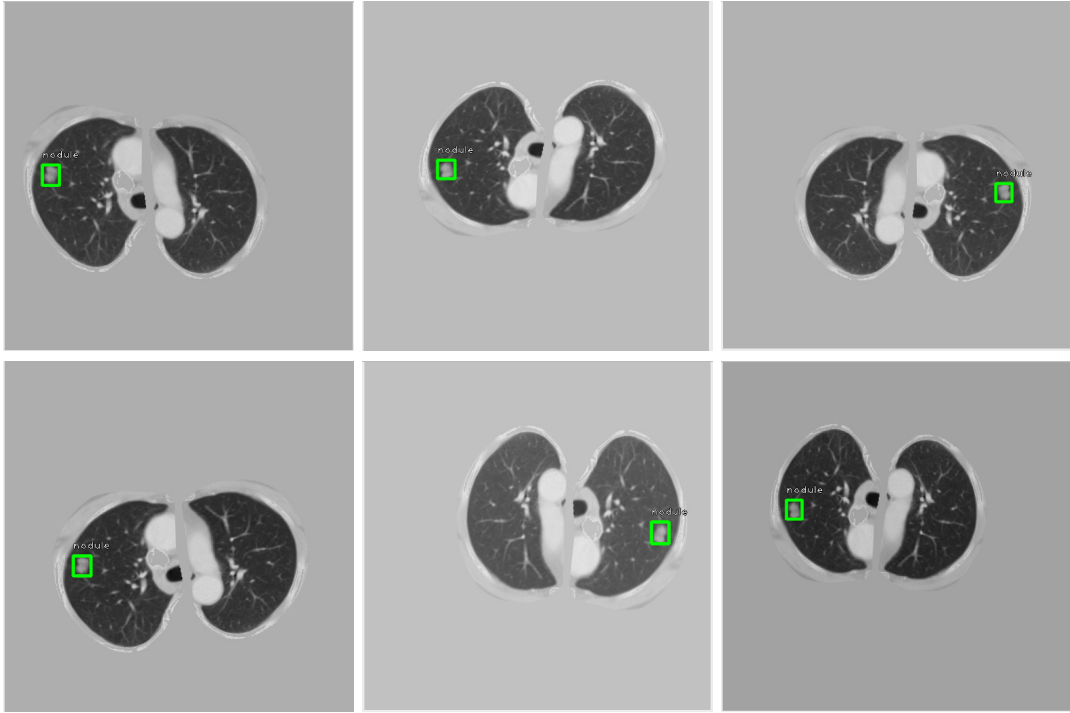


Figure 4: Image data augmentation examples. First row, from left to right: Original image, vertically flipping, horizontally flipping. Second row: Rotation with random degrees and random brightness/contrast adjustment. Green rectangles indicate nodule bounding box ground-truth.

representing the location and size of the nodule in the CT slice. We split the dataset for training, validation and testing using 80:1:1 ratio.

To avoid overfitting, we apply some data augmentation techniques to generate new samples from original training data. The techniques include using one lower and one upper slices of the ground-truth nodule, randomly rotating, vertical/horizontal flipping or shifting, or adjusting brightness/contrast from the input image. We apply image augmentation techniques only on the training dataset, and not to the validation/testing dataset. Fig. 4 shows examples of different augmented outputs from the same original image. Since a CT image usually includes not only the lung but also other tissues or vascular, we apply image segmentation proposed in [26] to remove the irrelevant background of CT images.

4.2. Experimental setup

Hyperparameter settings

We train the three detection models from scratch using the actual size of CT slices (512×512 pixels) as input and the original backbone networks as proposed by the authors: ResNet50 for Faster R-CNN and RetinaNet, and Darknet-53 for YOLOv3. All models were trained using Adam optimization, 10^{-5} of initial learning rate and 10^{-3} of clipnorm. Both the weight decay and bias decay were set to 5×10^{-4} . We use batch size of 2 due to VRAM limitation. The training was repeated for 100 epochs.

Table 1: Effectiveness of enriched data for LUNA16 dataset

Methods	Training set	Sensitivity (%) on test set		
		LUNA16	LUNA16+K	$\pm$
Faster R-CNN	LUNA16	83.05	80.08	-2.97
YOLOv3	LUNA16	85.60	84.61	-1.01
RetinaNet	LUNA16	88.13	86.92	-1.21
Faster R-CNN	LUNA16+K	84.74	85.38	0.64
YOLOv3	LUNA16+K	88.98	92.30	3.32
RetinaNet	LUNA16+K	91.52	95.38	3.86

Hardware configuration

Our experiments were performed using Python 3.7 on our private server equipped with an Intel Xeon E5-2620 v3 (2.4GHz) and a GeForce RTX 2080 Ti with 11GB GDDR6 256-bit memory.

4.3. Evaluation metrics

To compare with other methods in the same LUNA16 dataset, we use sensitivity as the main metric in our evaluation. Sensitivity, also known as the true positive rate, is calculated as the ratio between true positives over the total number of positive cases. In such metric, a detection is considered as true positive if the center location of the detection result is within the radius from the ground-truth nodule center. Outside this range, candidates are considered as false negatives. Therefore, sensitivity can be calculated as follows

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

where, TP is the number of true positives, FN is the number of false negatives.

For better comparison between different methods, the LUNA16 evaluation measures sensitivity at seven predefined false positive rates: 1/8, 1/4, 1/2, 1, 2, 4, and 8 false positives per scan. A Free Response Receiver Operating Characteristic (FROC) curve is a visual representation of these sensitivity values (on vertical axis) with its corresponding false positive rates (on horizontal axis). The average of these sensitivity values at seven false positive rates is defined as Competition Performance Metric (CPM) in the LUNA16 challenge evaluation.

4.4. Experimental results

In this section, we first evaluate the effectiveness of each proposed data-centric contribution. Then, we use combination of both contributions to compare with other works on the same LUNA16 test set. To ensure evaluation consistency, we include all images used in LUNA16 test set into LUNA16+K testset.

Effectiveness of enriched LUNA16 dataset

Firstly, we evaluate if additional data from the K dataset contributes to the overall detection performance. To do this, we train three detection model separately on LUNA16 and LUNA16+K datasets, then test on LUNA16 and LUNA16+K datasets.

Table 2: Effectiveness of customized anchor boxes on LUNA16 dataset

Methods	Test sensitivity (%)		
	Original	Customized	$\pm$
	Anchor boxes	Anchor boxes	
Faster R-CNN (ISODATA) [19]	91.4	-	-
Faster R-CNN [4]	89.1	-	-
Faster R-CNN [21]	86.4	-	-
YOLOv3 [18]	90.0	-	-
Faster R-CNN	83.05	87.29	4.24
YOLOv3	-	85.60	-
RetinaNet	88.13	88.98	0.85

Table 1 compares and summarizes detection performance of Faster R-CNN, YOLOv3, and RetinaNet on both LUNA16 and LUNA16+K datasets. The top half of this table shows performance of the three models trained on LUNA16, then tested separately on LUNA16 and LUNA16+K datasets. This scenario demonstrates the difficulties when models are trained without knowledge of large nodules: All models have lower sensitivity when tested on LUNA16+K than on LUNA16. The negative difference in this scenario is around 1.73%.

Once trained on LUNA16+K dataset, all three models perform better significantly, as illustrated on the bottom half of Table 1. When tested on LUNA16, all models gain around 2.8% sensitivity compared with the trained/tested on LUNA16 scenario, in which the most considerable improvement is 3.39% from RetinaNet (88.13% to 91.52%). When considering additional K dataset both for training and testing, the difference is most significant on RetinaNet at 7.25% (95.38% trained/tested on LUNA16+K and 88.13% trained/tested on LUNA16).

From this improvement, it can be inferred that RetinaNet is very sensitive to the dataset size, especially with the variety of object sizes and aspect ratios from the training set. In contrast, Faster R-CNN does not exhibit such a clear difference with additional data.

Effectiveness of customized anchor boxes

Secondly, we show the effectiveness of customized anchor boxes for popular object detection approaches. To do this, we train and evaluate each model with and without anchor box customization on the LUNA16 dataset, and compare the test results. We do not test the models on LUNA16+K dataset to be consistent with other state-of-the-art methods.

Table 2 summarizes detection performance of our customized anchor box sizes with Faster RCNN, YOLOv3 and RetinaNet on LUNA16 datasets. Overall, it can be seen that this customization yields better in detection performance than those with original anchor box sizes. Faster R-CNN benefits the most from this optimization, with 4.24% enhanced sensitivity (from 83.05% to 87.29%). On the other side, RetinaNet receives negligible improvement, from 88.13% to 88.98% (only 0.86% better), due to the fact that its default anchor box sizes are relatively small and compatible with small nodule sizes. This shortens the difference between Faster R-CNN and RetinaNet. It is worthy noting that since YOLOv3 applies anchor box customization by default using K-means clustering, there is no baseline to compare with. As a result, after customization, YOLOv3 performance is lower when compared with other two methods.

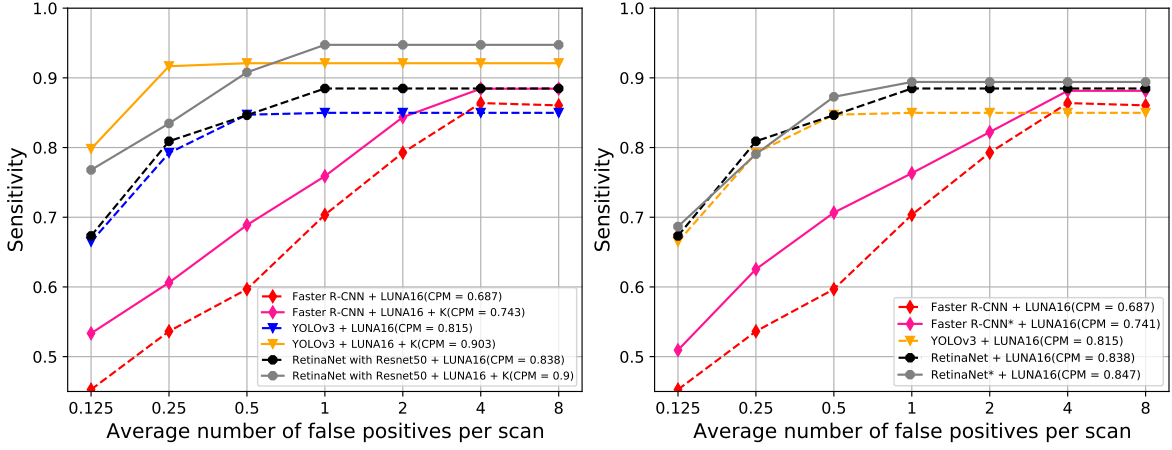


Figure 5: FROC curve comparison between LUNA16 and LUNA16+K datasets (left) and customization of anchor box sizes (right). Asterisks indicate methods having anchor box customization.

FROC evaluation

Fig. 5 compares FROC curves of the three methods on two different cases: with/without K dataset (left) and with/without anchor box customization (right). In this figure, methods having asterisk are those with anchor box customization. From these FROC curves, it can be depicted that RetinaNet outperforms the other two detection methods at high false positive rates, with both cases. This leading advantage starts from 0.5 FP/scan (with anchor box customization) and 1 FP/scan (with K dataset).

In contrast, Faster R-CNN improvement over different false positive rates is much lower than RetinaNet's and YOLOv3's. This can be explained as a result of being a two-stage method: Faster R-CNN not only requires more computational power for training and inferring, but also requires more data to train. This result indicates that Faster R-CNN can be useful in the cases in which a lot of training data are provided.

At very low false positive rates (1/8 and 1/4), YOLOv3 is among the best methods, especially in case when additional K dataset is provided, having sensitivity of 79.8% and 91.6%, respectively. It is also noticed that YOLOv3 benefits the most from the additional data from K dataset, especially at 1/8 and 1/4 FP rate. The difference is more compelling at 1/8 FP/s (13.3%) than at 1/4 FP/s (12.4%).

Comparison with other methods

Table 3 presents performance comparison of our method using additional K dataset and anchor box customization, with other methods in the literature on the same LUNA16 dataset. As it can be depicted from this table, by using our data-centric method, object detection frameworks can produce competitive results with the state-of-the-art studies. For example, YOLOv3 with K dataset can be top-4 in terms of sensitivity, but top-2 in terms of CPM, indicating a good detection network at all false positive rates as a whole. RetinaNet can benefit the most from our approach, being top-2 and top-3 for detection sensitivity and CPM, respectively.

Visualization of detection result

Fig. 6 shows some detection results on both LUNA16 and K datasets: Red rectangles

Table 3: Performance comparison of our method with other works in the literature when tested on the same LUNA16 dataset. Asterisks indicates methods having anchor box customization.

Methods	Training data	Year	Sensitivity (%)	CPM (%)
Ding et al. [4]	LUNA16	2017	94.6	89.1
Xie et al. [21]	LUNA16	2019	86.4	77.5
S4ND [27]	LUNA16	2018	97.2	93.1
FocalMix [28]	LUNA16	2020	N/A	89.2
Embedded Multi branch CNN [29]	LUNA16	2020	87.1	83.0
Multilevel contextual 3D CNNs [30]	LUNA16	2017	92.2	82.7
3D DCNN [31]	LUNA16	2019	94.6	N/A
Faster R-CNN*	LUNA16 + K	2022	89.2	74.3
YOLOv3	LUNA16 + K	2022	92.3	90.3
RetinaNet*	LUNA16 + K	2022	96.2	90

indicate true nodule predictions with the corresponding probability scores while green rectangles show ground-truth bounding box labels. It can be depicted from this figure that the detection models with customized anchor box sizes produce consistent results, even for small nodules in the early stage, with considerable accuracy.

5. CONCLUSION AND PERSPECTIVES

Early detection of pulmonary nodule is very important to diagnosis and treatment of the patient. Current deep learning pipelines usually focus on selecting state-of-the-art deep learning models and perform model hypertuning to achieve better pulmonary nodule detection performance. This method is called model-centric approach, which sometimes may miss important features of lung nodule dataset to train detection models.

In this work, we propose to use a data-centric approach, which takes into account the specific features of lung nodule dataset (nodule sizes and aspect ratios) to train deep learning detection models. We also enrich training data of lung nodules with data collected from a Vietnamese hospital. This contribution aligns with the current development trend of data-centric deep learning, image processing in computer vision. The evaluation of our proposed method is performed on the three widely used object detection models (Faster R-CNN, YOLOv3 and RetinaNet).

The experimental results show that when using nodule sizes and aspect ratios for anchor box customization, performance of these detection models can be improved up to 4.24% on Faster R-CNN, from 83.05% to 87.29%. Besides, with the use of additional data from K dataset, the performance is increased up to 7.25% with RetinaNet (from 88.13% to 95.38%). When compared with other works, the proposed method is among top 3 in terms of sensitivity.

There exist several directions to continue this work. Firstly, we can include bigger nodule sizes collected in the K dataset for evaluation (i.e. those larger than 30mm). Secondly, more experiments with other anchor-less methods can be performed for better analysis. Last but not least, more preprocessing methods can be applied to further enhance quality of the input CT scan images.

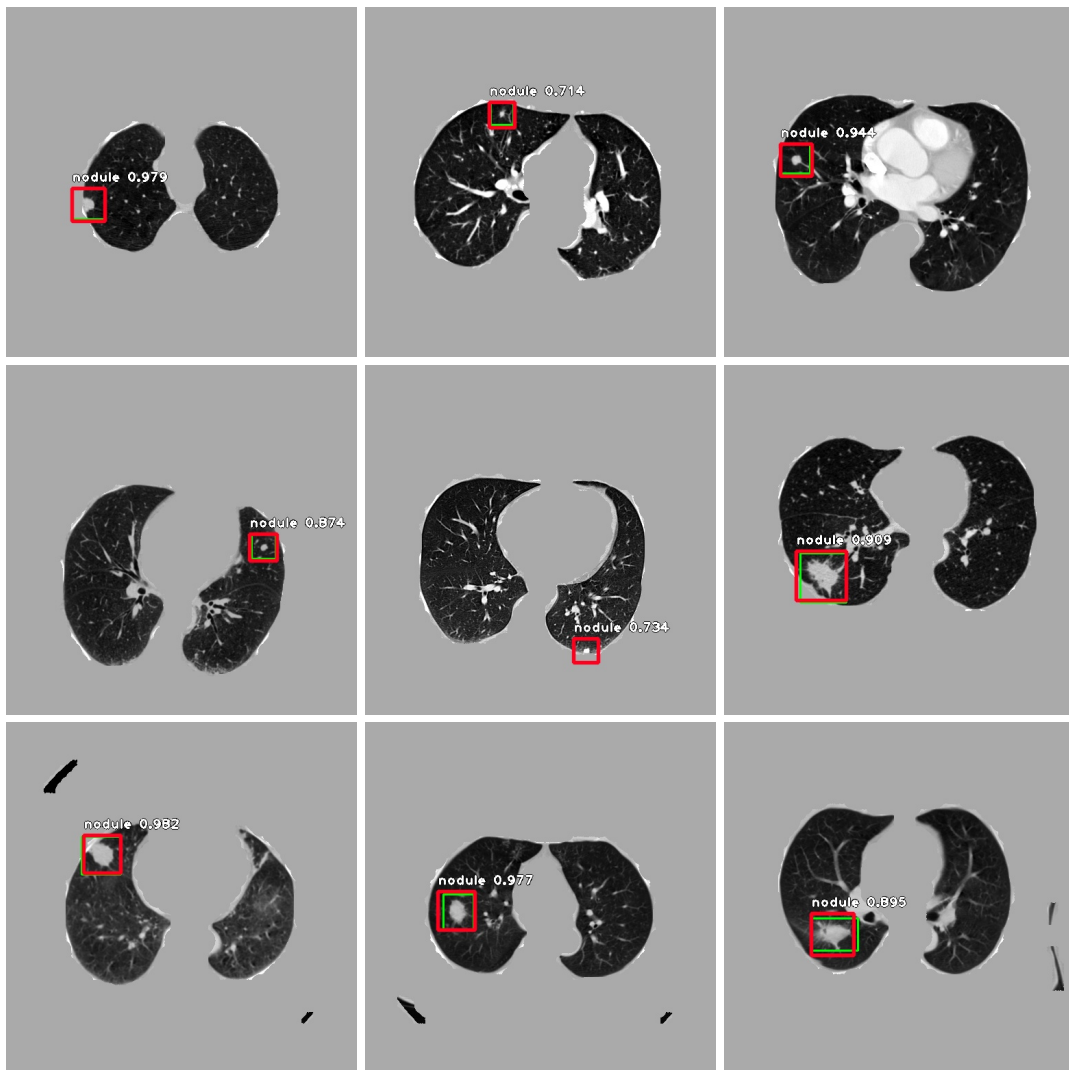


Figure 6: Examples of nodules detected by our method. Each true positive nodule is located at the center of the red rectangle. Green rectangles indicate ground-truth labels.

ACKNOWLEDGMENT

The authors would like to thank Vietnam Academy of Science and Technology (VAST) for funding this research in the scope of project DL0000.05/20-22. We also would like to thank Dr. Van Thi Nguyen and Dr. Duc Huu Duong of Vietnam National Cancer Hospital for their participation in the labelling and collection process of the K dataset.

REFERENCES

- [1] R. Siegel, K. Miller, H. Fuchs, and A. Jemal, “Cancer statistics, 2021,” *CA: A Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, 2021.
- [2] H. Sung, J. Ferlay, R. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers

- in 185 countries,” *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [3] A. C. Society, “Non-small cell lung cancer survival rates,” 2021. [Online]. Available: <https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/survival-rates.html>
 - [4] J. Ding, A. Li, Z. Hu, and L. Wang, “Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017. Lecture Notes in Computer Science*, vol. 10435. Springer, Cham, 2017.
 - [5] J. Wang, J. Wang, Y. Wen, H. Lu, T. Niu, J. Pan, and D. Qian, “Pulmonary nodule detection in volumetric chest ct scans using cnns-based nodule-size-adaptive detection and classification,” *IEEE Access*, vol. 7, pp. 46 033–46 044, 2019.
 - [6] Y. Li and Y. Fan, “Deepseed: 3d squeeze-and-excitation encoder-decoder convolutional neural networks for pulmonary nodule detection,” in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1866–1869.
 - [7] Z. Guo, L. Zhao, J. Yuan, and H. Yu, “Msanet multi-scale aggregation network integrating spatial and channel information for lung nodule detection,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2547–2558, 2022.
 - [8] A. Setio, A. Traverso, T. D. Bel, M. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. Fantacci, and B. G. 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, 2017.
 - [9] M. Motamedi, N. Sakharlykh, and T. Kaldewey, “A data-centric approach for training deep neural networks with less data,” 2021.
 - [10] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, 2015, pp. 91–99.
 - [11] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” 2018.
 - [12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
 - [13] S. A. III, G. McLennan, L. Bidaut, M. McNitt-Gray, C. Meyer, A. Reeves, B. Zhao, D. Aberle, C. Henschke, and E. H. et al., “The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans,” *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.
 - [14] W. Li, P. Cao, D. Zhao, and J. Wang, “Pulmonary nodule classification with deep convolutional neural networks on computed tomography images,” *Computational and Mathematical Methods in Medicine*, vol. 2016, p. Article ID 6215085, 2016.

- [15] J. George, S. Skaria, and V. V. et al., “Using yolo based deep learning network for real time detection and localization of lung nodules from low dose ct scans,” in *Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, p. 105751I.
- [16] W. Fan, H. Jiang, L. Ma, J. Gao, and H. Yang, “A modified faster r-cnn method to improve the performance of the pulmonary nodule detection,” in *Proc. SPIE 10806, Tenth International Conference on Digital Image Processing (ICDIP 2018)*, 2018, p. 108065A.
- [17] I. W. Harsono, S. Liawatimena, and T. W. Cenggoro, “Lung nodule detection and classification from thorax ct-scan using retinanet with transfer learning,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 3, pp. 567–577, 2022.
- [18] C. Liu, S.-C. Hu, C. Wang, K. Lafata, and F.-F. Yin, “Automatic detection of pulmonary nodules on ct images with yolov3: Development and evaluation using simulated and patient data,” *Quantitative Imaging in Medicine and Surgery*, vol. 10, no. 10, 2020.
- [19] C. Tong, B. Liang, M. Zhang, R. Chen, A. K. Sangaiah, Z. Zheng, T. Wan, C. Yue, and X. Yang, “Pulmonary nodule detection based on isodata-improved faster rcnn and 3d-cnn with focal loss,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 1s, pp. 1–9, 2020.
- [20] A. Traoré, A. O. Ly, and M. A. Akhloufi, “Evaluating deep learning algorithms in pulmonary nodule detection,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 1335–1338.
- [21] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, “Automated pulmonary nodule detection in ct images using deep convolutional neural networks,” *Pattern Recognition*, vol. 85, pp. 109–119, 2019.
- [22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.
- [23] J. Redmon and A. Farhadi, “Yolo9000: Better, faster, stronger,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263–7271.
- [24] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [26] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, “Evaluate the malignancy of pulmonary nodules using the 3d deep leaky noisy-or network,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3484–3495, Nov 2019.
- [27] N. Khosravan and U. Bagci, “S4nd: Single-shot single-scale lung nodule detection,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, ser. Lecture Notes in Computer Science. Springer, Cham, 2018, vol. 11071.

- [28] D. Wang, Y. Zhang, K. Zhang, and L. Wang, “Focalmix: Semi-supervised learning for 3d medical image detection,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3950–3959.
- [29] W. Zuo, F. Zhou, and Y. He, “An embedded multi-branch 3d convolution neural network for false positive reduction in lung nodule detection,” *Journal of Digital Imaging*, vol. 33, no. 4, pp. 846–857, 2020.
- [30] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, “Multilevel contextual 3d cnns for false positive reduction in pulmonary nodule detection,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2016.
- [31] A. Masood, P. Yang, B. Sheng, H. Li, P. Li, J. Qin, V. Lanfranchi, J. Kim, and D. D. Feng, “Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest ct,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–13, 2019.

Received on June 13, 2022

Accepted on September 05, 2022